# Fleet Management Strategies for Shared Automated Vehicles in Different City Scales

To obtain the degree of Master of Science at Delft University of Technology,
to be defended publicly on 17th July, 2025 at 10:00 AM.

Heran Zhao
Civil Engineering and Geosciences, TU Delft

**Project duration:** February, 2025 – July, 2025

**Graduation Committee:**

| | |
|---|---|
| Dr.ir. I. Martínez, Daily Supervisor | CEG, TU Delft |
| Dr. B. Atasoy, Chair and Supervisor | ME, TU Delft |
| Prof.dr. O. Cats, External Committee Member | CEG, TU Delft |

**Report Number:** 2025.TIL.9083

**Abstract**

Shared Automated Vehicles (SAVs) hold significant potential to redefine urban mobility services. However, the applicability and optimization of their operational strategies in diverse urban contexts remain unclear, particularly concerning the complex interactions with existing public transport systems. This study systematically investigates how urban heterogeneity, including city scale, network topology, and demand patterns, modulates SAV fleet management strategies to balance operator costs and user utility within a multi-modal transportation context.

To this end, this study develops a comprehensive simulation framework that combines an advanced ride-pooling candidate generation algorithm (ExMAS) with an innovative two-stage vehicle assignment optimization process, and integrates a nested Logit model to quantify the competition structure with public transport. This framework is applied to 37 Dutch cities of varying scales to derive generalizable findings.

The study finds that population scale is a fundamental determinant of the required fleet size, exhibiting a strong linear relationship, especially in large cities. The key to enhancing operational efficiency, however, lies in more nuanced urban structure metrics. In large cities, longer commuting distances combined with more complex networks (higher node degrees) foster ride-pooling potential. In contrast, for small cities, the local compactness of the network (higher clustering coefficient) and demand density are critical for improving vehicle turnover efficiency. Furthermore, the research confirms that a uniform pricing strategies are unlikely to achieve best performance on pooling efficiency, highlighting the necessity of implementing differentiated pricing based on user preferences and city scale. In competition with public transport, the study's *Public Transport Competitiveness Index* reveals a non-monotonic relationship with travel distance. A state of competitive balance is observed for short-distance trips, while PT holds a distinct advantage in the medium-distance range. For long-distance trips, the inherent speed advantage of SAVs allows them to become the more competitive option. This external competition also structurally increases the internal pooling rate of the SAV system.

The findings of this study provide critical strategic insights for SAV operators and policymakers, emphasizing the critical importance of adjusting fleet management, service design, and pricing strategies according to specific urban characteristics and the competitive environment, thereby providing decision support for achieving an efficient and sustainable urban transportation system.

**Keywords:** Shared Automated Vehicles, Fleet Management Strategies, Multi-City Analysis, Ride-Pooling, Multi-Modal Transportation

# Contents

# Chapter 1

# Introduction and Motivation

The rise of automated vehicles (AVs) is becoming an irreversible trend, which not only has provided many industrial logistics and shipping solutions but also has already created new possibilities in urban mobility services , particularly through the integration of shared automated vehicles (SAVs) into existing transport systems (Salazar et al., 2018).The introduction of SAVs will undoubtedly have a great impact on urban transportation. For example, Liu et al. (2024) points out that by introducing SAVs, cities can effectively mitigate congestion, boost compact urban development, leading to a 14.0% increase in regional accessibility and a 18.5% decrease in total travel time. Furthermore, a case study in Tokyo shows that the introduction of SAVs caused 14%-32% of commuters to shift from traditional modes to SAVs, but an excessive introduction of SAVs may lead to a decrease in the number of people using transportation methods with minimal environmental impact (like walking or cycling), thereby affecting health and the environment (Ishibashi and Akiyama, 2022). Moreover, as a supplement to public transport, the interaction between SAVs and the existing public transportation system also needs to be emphasized. Cats et al. (2022) found that ride-hailing services both compete with and complement public transport: they fill accessibility gaps in under-served areas but compete for demand in well-served regions.In areas with high SAV supply, the introduction of SAVs may effectively reduce reliance on private cars, while in cities that heavily depend on public transport, the emergence of SAVs may negatively impact the modal share of buses(Tak et al., 2021).Therefore, for SAV operators, determining the operation strategies has become the key to the process of implementing various services related to SAVs.

Ride-pooling, as a core mechanism for enhancing the efficiency and sustainability of SAV systems, has been widely demonstrated by numerous studies to significantly reduce the required number of vehicles and total vehicle kilometers traveled (VKT), thereby alleviating urban congestion and emissions (Fagnant, 2015; Balac et al., 2020; Alonso-Mora et al., 2017). At the same time, it improves vehicle utilization, reduces operating costs, and shortens user waiting times (Jin et al., 2021; Militão and Tirachini, 2021). In addition, ride-pooling plays an important role in improving the sustainability of urban transport systems, reducing carbon emissions, and lowering energy consumption (Fagnant, 2018). However, the efficiency of ride-pooling is highly dependent on multiple factors such as urban structure, demand distribution, and operational strategies, the potential and system benefits of ride-pooling vary significantly under different city and demand patterns(Fagnant, 2018; Soza-Parra et al., 2024; Alonso-Mora et al., 2017). Therefore, systematically studying the role and optimization paths of ride-pooling mechanisms in the context of multiple cities and diverse demand structures is a key prerequisite for achieving efficient and sustainable operation of SAV system

The core challenge in achieving this key prerequisite lies in the current lack of practical operational experience with SAVs (Narayanan et al., 2020). This lack of experience means that operators and policymakers may not be able to determine whether SAVs can achieve similar effects in cities of different scales, especially given that urban form and travel demand patterns significantly influence ride-pooling potential and overall system efficiency. For example, Zwick et al. (2021) conducted a study on six real communities within the Munich metropolitan area in Germany, each with different population sizes and travel densities (representing various city scale indicators). The results revealed a logarithmic growth relationship between SAV system efficiency (such as the ratio of passenger-booked kilometers to vehicle kilometers traveled) and travel density. Similarly, Soza-Parra et al. (2024), taking Amsterdam in the Netherlands as a case, systematically varied parameters such as the number of attraction centers, the dispersion of destinations around each center, and the distribution of trip lengths. They found that the number of attraction centers, destination density, and trip distance distribution all have a decisive

impact on ride-pooling efficiency. The study also showed that ride-pooling efficiency varies with city scale and travel density—performing better in scenarios with longer travel distances and more concentrated demand. The shift from monocentric to polycentric demand distribution patterns has limited impact on efficiency, but their analysis is limited to the demand characteristics of a single city. In addition, the interrelationship between service demand, achievable service levels (such as passenger waiting time), and required fleet size also varies significantly with city characteristics. For example, Boesch et al. (2016) found in their study of the Greater Zurich area that service performance drops significantly at low demand levels, and that the maximum acceptable waiting time for passengers is a key factor in determining the required fleet size. These differences highlight a key challenge: strategies effective in one urban context may not be directly applicable to another. Empirical studies also suggest that in scenarios with low ride-pooling efficiency, the overall benefits of SAV systems may fall short of expectations (Kumakoshi et al., 2021; Jin et al., 2021), further underscoring the importance of optimizing operational strategies according to specific conditions and forcing operators to carefully consider these external constraints in decisions such as pricing discounts, service level commitments, and fleet size planning.

The efficient operation of SAV systems depends not only on the optimization of ride-pooling mechanisms, but is also influenced by multiple factors such as urban structure, demand distribution, and interactions with public transportation. Therefore, this study aims to explore how demand characteristics, transport network structure, and competition with public transit jointly affect the performance of SAV systems under different city scales. Based on an understanding of these mechanisms, this research is dedicated to providing strategic-level insights for operators in formulating fleet management strategies for SAVs.

## 1.1 Research Scope

To ensure that this study can accurately and effectively address the aforementioned research questions, the following sections will first clearly define the core elements and basic assumptions of the SAV operational scenarios considered in this research, and then elaborate on the specific connotation and analytical framework of 'fleet management strategies' in this study.

Before discussing specific fleet management strategies, it is necessary to clarify the operational characteristics and research perspective of SAVs as discussed in this study. According to the multi-dimensional classification framework of SAV services by Narayanan et al. (2020), the SAVs in this research refer specifically to systems that are on-demand, allow users to choose between private and shared rides, are centrally owned and managed by operators (independent system), and adopt a fixed pricing strategy. Under this setting, the study further focuses on static demand scenarios (the entire simulation period is one day) for analysis. This is mainly because, from the perspective of SAV operators, fleet management of SAVs is essentially a strategic planning problem (Fan et al., 2023). Analyzing within the framework of static demand not only strips away the dynamic complexity of real-time operations and allows the analysis to focus on key strategic parameters, but also helps to fairly compare different strategies in standardized scenarios and ensure the repeatability of results, thereby better assisting operators in evaluating the long-term strategic deployment of SAVs. Furthermore, clarifying the key differences between SAVs and human-driven vehicles from the operator's perspective is also crucial for understanding subsequent fleet management strategies. For example, in terms of vehicle dispatching and control, Ashkrof et al. (2020) points out that drivers of human-driven vehicles may choose whether to accept orders based on personal criteria, while the order-matching process for SAVs is entirely controlled by a central operating system. Regarding the composition of operating costs, Bösch et al. (2018) mentions that although SAV systems do not bear driver-related labor costs, they require higher upfront technological investment and ongoing vehicle cleaning and maintenance expenses.

Based on the above classification of SAV services and their differences from human-driven vehicles, the following settings are considered in this study:

- Consider a duration of 24 hours for travel demand
- Do not consider cancellations from SAVs side
- Do not consider driver cost

In order to better reflect real-world conditions, several simplifying assumptions were made in the modeling process of this study. The available travel modes are limited to public transport and SAVs, with SAVs providing both private and pooled ride services. Therefore, travel demand will be allocated

between these two modes. For public transport, only walking distance is considered as the basis for utility calculation, and walking is not treated as a separate travel mode.

On this basis, the analytical perspective and core objectives of "fleet management strategies" in this study are further clarified. The research on fleet management strategies for SAV systems in this study does not aim to determine a single optimal fleet size, but rather focuses on how operational decisions and external factors affect overall system performance. The main objective is to minimize the joint utility cost of the SAV system, that is, to combine operator and user costs through a weighted cost function, while ensuring that 100% of travel demand is met. In this study, a successful SAV system should not only pursue its own operational efficiency but also maintain basic attractiveness to users (Wen et al., 2018). Therefore, the minimization of total cost is not intended to precisely simulate the real-time decision-making behavior of operators or users, but rather to abstract the strategic balance between cost and service as a mathematical optimization objective. The trade-off between operator and user costs is achieved by adjusting the weights of each cost component in the joint utility cost function. This study evaluates the impact of changes in several key operational parameters on system response, including ride-pooling discounts, maximum pickup waiting time (related to service level commitments), and the "willingness-to-share resistance factor" (WtSR) reflecting user preferences. In addition, external environmental characteristics are also included in the analysis, such as city scale (including population, road network structure, and demand patterns) and its competition with public transport.

## 1.2 Research Questions

The scope of this study establishes a framework for analyzing SAV fleet management, focusing on the trade-offs between operator and user costs under specific operational parameters and external conditions. To translate this framework into a structured inquiry, this study poses a series of research questions designed to systematically explore these complex interactions and their implications for SAV operators:

In cities of varying scales, what adaptations should be made to SAV fleet management strategies to effectively balance operator and user costs, while considering the competitive interactions with public transport services?

**Sub-questions:**

1. What are the relationships between specific city characteristics (e.g., network topology, demand patterns) and key SAV performance indicators, and how do these relationships differ across city scales?

2. What are the interaction effects between passenger willingness-to-share and ride-pooling fare discounts, and what are the implications for designing targeted, preference-aware pricing strategies across different city scales?

3. What are the scaling effects of increasing demand volume on SAV system performance, and what is the resulting shift in the influence of passenger preferences on ride-pooling behavior across different demand scales?

4. What is the method to extend the ExMAS framework to capture the competitive relationships between SAVs and public transport?

5. What shifts occur in the competitive structure between SAVs and public transport in response to PT fare discounts and traffic congestion, and what are the resulting impacts on SAV system performance?

## 1.3 Research Overview

To address the research questions, this study employs a quantitative approach centered around simulation, mathematical modeling, and statistical analysis. This integrated methodology allows for a comprehensive exploration of how operational parameters and external factors affect key performance indicators (KPIs), including joint utility cost, operational efficiency (e.g., vehicle utilization, extra mileage), service level (e.g., passenger waiting time), and the required fleet size. The trade-off between operator and user costs is central to the optimization process, as the operator's economic viability is directly linked to its

ability to attract and retain users by offering a competitive level of service. By systematically analyzing the simulation outputs, the goal is to provide simulation-driven insights that enable SAV operators to refine fleet management strategies and assist policymakers in designing effective subsidy schemes, ensuring that service efficiency is aligned with broader urban mobility objectives across different city scales.

The main contributions of this thesis are as follows:

- **Systematic analysis of multi-city heterogeneity**: This study utilizes real-world data from multiple cities to systematically compare the applicability and optimization pathways of SAV fleet management strategies under different city scales, network structures, and demand patterns, thereby enhancing the generalizability and practical relevance of the findings.
- **Explicit modeling and strategic analysis of competition between SAVs and public transport**: This study explicitly characterizes the market share and interaction between SAVs and public transport in different urban environments, providing a theoretical foundation for the joint optimization of multimodal transportation systems.
- **Integrated analytical framework for ride-pooling scheme generation and executable dispatch**: This study develops a two-stage integrated approach: the first stage adopts an event-driven vehicle assignment mechanism to explicitly construct the decision space for vehicle-passenger relationships, ensuring that all dispatch schemes are conflict-free in terms of time and vehicle resources; the second stage further determines the optimal vehicle dispatch and passenger assignment scheme through global optimization. This approach bridges the gap between theoretical analysis and practical operations.

The remainder of this thesis is structured as follows. Chapter 2 provides a comprehensive literature review on SAV fleet management and related topics. Following this, Chapter 3 details the methodological framework developed for this study, including the extension of the ExMAS platform to incorporate multimodal choice models, which directly addresses research question 4. Chapter 4 then describes the simulation scenarios, parameter settings, and key performance indicators used for the analysis. The core findings are presented in Chapter 5, which investigates the impact of urban characteristics (addressing research question 1), analyzes key operational parameters (addressing research questions 2 and 3), and examines the competitive relationships with public transport (addressing research question 5). Following this, Chapter 6 synthesizes the findings into practical recommendations for operators and critically discusses the study's limitations. Finally, Chapter 7 summarizes the study's main contributions and key findings in direct response to the research questions, and suggests directions for future research.

# Chapter 2

# Literature Review

This Chapter conducts a state-of-art analysis to explore the topics, main methodologies, research gaps in SAV fleet management and operations, to illustrate the importance of SAV fleet management on different city scales.

Key questions addressed in this section include:

- What are the relevant topics of SAV fleet management?
- What are the main stems of methodologies used for addressing SAV fleet management challenges?
- What are the research gaps in the SAV fleet management field?

By answering these questions, this section will illustrate the connections and differences between the collected articles. The following content of this section will be divided into three parts. First, an analysis of the application topics, followed by a conclusion of the methodology, the last part is the discussion about topics and the future gaps.

## 2.1 Research Topics Related to SAVs

To clarify the positioning and contribution of this study in the field of SAV operations, this section provides a broad review of relevant research based on the scope of this study, with a focus on two core themes: SAV fleet management and operational strategies, and the integration of SAVs with existing transport systems. Reviewing the main topics of these cutting-edge studies is crucial for illustrating the subject and research value of this work.

### 2.1.1 SAV Operational Strategies

Hyland and Mahmassani (2017) identified 17 categories of SAV management topics, including fleet size flexibility, reservation structures, and pricing mechanisms. In the area of SAV fleet management and operational strategies, researchers have explored the influence of various factors. For example, Wang et al. (2018) addressed fleet configuration problems through route optimization and validated their algorithm using the MATSim framework, demonstrating its effectiveness in dynamic dispatching, minimizing empty mileage, and improving occupancy rates, thus providing quantitative deployment solutions for urban traffic managers.

Research exploring SAV deployment from both strategic and operational perspectives includes Seo and Asakura (2022), who proposed the broad impacts of SAVs in future cities. Monteiro et al. (2021) studied the optimization of fleet size under round-trip and one-way modes.

Regarding different service level strategies, Militão and Tirachini (2021) analyzed the impact of rejection strategies, vehicle capacity, and other factors on service levels (such as average waiting time and detour factor), especially under zero-rejection or constrained operations. Schröder and Kaspi (2024), in a Munich case study, explored how to integrate spatially varying external costs (such as air pollution, noise, accidents, and climate impacts) into the operational decisions of automated ride-pooling fleets to optimize their social benefits. Other studies on operational strategy parameters have also focused on whether to allow passenger requests to be rejected (Militão and Tirachini, 2021) and whether to support ride-pooling services (Seo and Asakura, 2022).

In addition, changes in demand levels also have a profound impact on the efficiency of SAV systems. For example, Boesch et al. (2016) found in their study of the Greater Zurich area that service performance drops significantly at low demand levels, and that the acceptable passenger waiting time is a key factor in determining fleet size. Jin et al. (2021) explored the impact of setting an upper limit on fleet size, while Vosooghi et al. (2019) evaluated the effects of different SAV fleet sizes, vehicle capacities, and operational strategies (ride-pooling, vehicle rebalancing) on the transport system in the Rouen metropolitan area of France.

As an important research topic in shared mobility, the impact of ride-pooling has received widespread attention. Balac et al. (2020) studied the reduction of required fleet size by integrating passenger demand in time and space and using vehicles of different capacities, finding that ride-pooling can significantly reduce the number of vehicles needed. Kucharski and Cats (2020) investigated the relationship between ride-pooling and profitability, finding that the profitability of ride-pooling depends on the balance between fare discounts and operational cost savings (e.g., reduced vehicle mileage). For example, discounts of 10% to 30% can generally lead to profitability, but higher discounts may result in reduced profitability.

### 2.1.2 Integration of SAVs with Existing Transport Modes

In addition, some studies focus on the integration and impact of SAVs with existing urban transport systems, especially in the context of multimodal travel and mixed traffic flows. For example, regarding the impact of SAVs on urban traffic, Jin et al. (2021) focused on the effect of shared mobility systems on traffic congestion and analyzed the optimal density of SAVs as for-hire vehicles (FHV) in scenarios with mixed traffic flows including private cars and SAVs. In mixed systems (including both human-driven vehicles and SAVs), Militão and Tirachini (2021) systematically analyzed the effects of fleet size, rejection strategies, vehicle capacity, and other factors on service levels and system efficiency. There are also studies examining how combinations of different travel modes can meet urban demand, such as Fan et al. (2023), who investigated how only two travel modes (SAVs and bicycles) can satisfy urban travel needs. Vosooghi et al. (2019) focused on the impact of changes in fleet size and passenger capacity on mode choice. Other studies have taken a more macro perspective to explore the potential impact of SAV introduction on VKT (vehicle kilometers traveled). For example, Fielbaum and Pudāne (2024) pointed out that in public transport-oriented cities, SAVs may significantly increase VKT, and studies such as Fagnant (2015, 2018) also support the view that the introduction of SAVs may lead to an increase in VKT.

It can be said that, the existing literature has systematically explored SAV operations, covering various aspects such as integration with existing transport systems, fleet management and operational strategies, ride-pooling mechanisms, and demand characteristics and urban environments, demonstrating a rich diversity of research perspectives and methods. Different studies have their own emphases in theoretical modeling, simulation analysis, and practical applications, providing a solid foundation for understanding the complexity and diversity of SAV systems. It should be emphasized that demand characteristics and urban heterogeneity are key factors profoundly affecting the performance of SAV systems. As detailed in the introduction (Section 1), these factors specifically include city scale, demand density, and trip distance, among others.

In addition, this thesis, when reviewing relevant research, particularly focuses on two directions closely related to model construction: operational strategies and multimodal integration. The core purpose is to clarify how existing literature defines and handles key variables, thereby providing theoretical support for parameter selection and modeling. On the one hand, in terms of fleet configuration and dispatch mechanisms, existing studies offer rich experience regarding key variables such as service level (e.g., maximum waiting time), operational efficiency (e.g., empty mileage rate, occupancy rate), and ride-pooling incentive mechanisms (e.g., discount strategies). These studies not only provide a rationale for the operational parameters examined in this thesis, but also offer important theoretical references and empirical foundations for constructing a modeling framework that comprehensively considers user utility and operational costs.

On the other hand, regarding the competitive relationship in multimodal integration and differences in urban structure, although existing literature has discussed these issues, most lack explanatory mechanisms or unified indicator settings, especially regarding how PT pricing, network structure, and demand density affect SAV performance. In this competitive environment, public transport pricing strategies are no longer isolated decisions, but need to be strategically considered in relation to the service level and user attractiveness of SAVs, in order to maintain the service value and market share of public transport, thereby affecting the balance and efficiency of the entire transport system. Therefore, the studies

discussed in this section not only provide necessary parameter sources for modeling, but also reveal the shortcomings of existing research in variable selection, mechanism modeling, and scope of applicability, providing a theoretical foundation and room for improvement for this thesis to construct a scale-sensitive and competition-structured SAV operation model. Next, we will review the main methodologies adopted in academic research on SAV fleet management and further analyze the strengths and limitaions of mainstream research methods in this field.

Table 2.1 summarizes the main findings and conclusions in SAV operational strategy research.

Table 2.1: Summary of Main Findings in SAV Operational Strategy Research

| **Reference** | **Main Findings** |
|---|---|
| (Jin et al., 2021) | Fleet density can be derived based on critical density, and strategies for limiting fleet size are proposed to alleviate congestion caused by empty trips and detours. |
| (Qu et al., 2022) | The algorithm is computationally efficient, does not depend on the initial positions of vehicles, and allowing ride-pooling can reduce the required SAV fleet size by approximately 30%. |
| (Wang et al., 2018) | By adopting appropriate fleet size and deployment strategies, it is possible to meet demand while reducing the number of vehicles and increasing occupancy rates. |
| (Seo and Asakura, 2022) | Moderate introduction of ride-pooling can continuously improve the combined utility of passengers, operators, and society. |
| (Monteiro et al., 2021) | The flexibility of passengers' willingness to walk significantly affects service capacity. |
| (Fan et al., 2023) | The initial distribution of SAVs is influenced by population distribution, land use, and residents' travel behavior, and the study successfully addresses non-convex and nonlinear computational challenges. |
| (Balac et al., 2020) | Ride-pooling can reduce the required number of vehicles by up to 96% and also decrease vehicle kilometers traveled (VKT). |
| (Militão and Tirachini, 2021) | With zero rejection, average waiting time and detour factor are significantly positively correlated with vehicle capacity; under constrained operations, the relationship between fleet size and demand determines the rejection rate, and there are clear economies of scale. |
| (Vosooghi et al., 2019) | As fleet size increases, the proportion of SAV trips rises, but the total travel distance also increases. |
| (Boesch et al., 2016) | At low demand levels, service performance declines significantly; when demand reaches 5%-6%, fleet utilization decreases as demand increases; acceptable passenger waiting time is a key factor in determining fleet size. |

## 2.2 Relevant Methodologies for SAV Fleet Management Strategies

Mainstream approaches to SAV fleet management can be categorized into two groups: simulation-based methods and mathematical modeling methods. To effectively address the complex issues of SAV fleet management and policy analysis across multiple city scales and multimodal interactions, it is essential

to understand the strengths and limitations of these methodologies.

Simulation-based methods are capable of capturing the dynamic characteristics, multi-agent interactions, and stochasticity of transportation systems in detail, thus becoming a crucial tool in SAV research. In this field, researchers have primarily adopted the MATSim (multi-agent transport simulation) framework(Horni et al., 2016). For example, Vosooghi et al. (2019) utilized the dynamic vehicle routing problem (DVRP) extension in MATSim, integrating user preference changes into the model and employing insertion heuristics to determine AV routes for addressing the fleet sizing problem. Wang et al. (2018) combined MATSim with the demand-responsive transport (DRT) module to perform dynamic routing and passenger scheduling, then used MATSim's evolutionary algorithm to optimize fleet size and deployment strategies. In addition, Boesch et al. (2016) developed a proprietary simulation system to model AV behavior, using demand generated by MATSim as input for their simulation model. Schröder and Kaspi (2024) further improved the agent-based simulation tool FleetPy (Engelhardt et al., 2022), incorporating external costs as the main optimization objective to guide routing and ride-sharing decisions, and demonstrated the potential for application in different urban environments due to its modularity. Zwick et al. (2022) discussed mainstream simulation methods, noting that although mobiTopp/FleetPy and MATSim share similar goals in simulating shared mobility systems (such as in the case of Hamburg), their methods and implementation paths differ significantly, reflecting the specialization and potential complementarity of different simulation tools in addressing specific problems: mobiTopp/FleetPy is suitable for rapid simulation of small-scale, real-time shared mobility operations, while MATSim is more appropriate for analyzing the potential impacts of large-scale shared mobility services on the overall transportation network and the long-term evolution of multi-agent behaviors.

Despite providing a general framework for analyzing SAV systems, the above simulation platforms exhibit limitations in specific operational aspects, such as efficient and attractive ride-sharing matching, strategic fleet management, and flexible control of operational parameters.

Therefore, for problems that require precise optimization and flexible parameter control, mathematical modeling methods demonstrate unique advantages. In this regard, some studies have adopted multi-objective optimization approaches to address the fleet size problem. For instance, Seo and Asakura (2022) proposed a multi-objective optimization model that divides SAV planning into strategic and operational levels, systematically analyzing issues such as fleet size, operational strategies, and scheduling optimization. Monteiro et al. (2021) employed a mixed-integer linear programming (MILP) model to optimize fleet size under one-way and round-trip modes, using MATSim for parameter calibration and validation. Fan et al. (2023) constructed a mixed-integer nonlinear programming model that comprehensively considers congestion effects and mode choice, subsequently applying the outer-inner approximation method and breakpoint generation algorithm to address the model's nonlinearity and non-convexity. Qu et al. (2022) transformed the fleet size problem into a minimum path cover problem in graph theory and solved it using the Hopcroft-Karp algorithm(Boesch and Gimpel, 1977). Balac et al. (2020) formulated their research as a MILP model, a variant of the discrete minimum cost flow problem, and validated it using the EQUASIM framework supported by MATSim(Hörl and Balac, 2021). Militão and Tirachini (2021) proposed a total cost minimization model that covers both operating costs and user costs, using MATSim to generate model inputs and calibrate parameters. Jin et al. (2021) adopted a compartmental model that divides SAVs into four stages, using an extended bathtub model and point queue model to describe the dynamics within each stage. These mathematical programming models have advantages in precisely solving for the optimal solution under specific objectives, but when facing large-scale, dynamically changing urban transportation systems and the complex spatiotemporal constraints and multi-agent interactions in SAV operations, they often encounter excessive computational complexity, difficulty in capturing all operational details (such as specific vehicle sequencing and conflict avoidance), or excessive simplification of real-world scenarios. Particularly for this study, which requires detailed evaluation of the impacts of different operational parameters and policies in diverse urban environments, relying solely on traditional mathematical models is insufficient to fully achieve the intended objectives.

In the field of shared mobility, methods to improve ride-sharing efficiency have attracted widespread attention(Xia et al., 2015; Kucharski and Cats, 2020; Agatz et al., 2012, 2011; Furuhata et al., 2013). However, real-time simulation platforms for generating ride-sharing solutions mostly rely on heuristic or rule-based approaches, making it difficult to systematically enumerate all possible combinations. As a mathematical optimization method, Kucharski and Cats (2020) proposed the ExMAS (Exact Matching of Attractive Shared-rides) algorithm, which is a utility-based approach for generating ride-sharing solutions by evaluating factors such as fare discounts, travel delays, and passenger discomfort to produce optimal ride-sharing plans. ExMAS represents the ride-sharing problem as a directed multigraph, reducing computational complexity through layered search and star expansion, thereby enabling sys-

tematic evaluation of operational parameters (such as ride-sharing discounts) and service design. This makes it particularly suitable for strategic fleet management under different demand patterns and city scales. Compared to other ride-sharing methods (such as time window-based approaches (Alonso-Mora et al., 2017)), ExMAS focuses more on long-term strategic optimization rather than real-time matching, providing a robust analytical framework for fleet size assessment and its integration with public transport systems. Table 2.2 summarizes the key differences between ExMAS and time window methods, highlighting the suitability of ExMAS for the objectives of this study.

Table 2.2: Comparison between the ExMAS algorithm and the time window-based method

| **Comparison Dimension** | **ExMAS** | **Time Window-based** |
|---|---|---|
| **Key Features** | Utility-based static demand optimization; provides a global, system-level perspective to support strategic decision-making. | Real-time matching of dynamic demand; focuses on short-term operational adjustments based on preset batch processing times. |
| **Advantages** | Supports systematic evaluation of the impact of different operational strategies by analyzing long-term demand patterns and optimizing the combined utility of passengers and operators. | Responds quickly to real-time demand changes, ensuring instant matching within the preset time window. |
| **Applicability for Strategy Evaluation** | The framework is suitable for evaluating system performance under different strategies (including cost, efficiency, user satisfaction, and the resulting resource requirements such as fleet size), assisting strategic decision-making, rather than being directly designed to output a single optimal fleet size. | Not designed for strategic evaluation; typically operates under a fixed fleet size to maximize short-term vehicle utilization. |
| **Matching Method** | Focuses on travel utility, dynamically balancing waiting time, detours, and system efficiency; suitable for static, reservation-based demand scenarios. | Matches based on strict time window constraints (e.g., 5 minutes), which may result in high levels of sharing with insufficient attractiveness. |
| **Application Scenarios** | Suitable for static demand analysis where all trips are known in advance, aiming to evaluate the impact of long-term strategies. | Most suitable for real-time operational scenarios requiring rapid response and instant matching. |

The ExMAS algorithm focuses on utility-based optimization and, as an open-source tool, provides a valuable foundation for generating potential ride-pooling solutions. However, although ExMAS performs well in identifying highly attractive, passenger-oriented ride-pooling options, its core functionality does not lie in generating vehicle-specific, practically executable, and conflict-free vehicle schedules. Instead, it estimates the required fleet size by analyzing peak concurrent trips. The algorithm is more concerned with evaluating theoretical resource requirements from the perspective of passenger matching and pooling potential, rather than directly outputting explicit vehicle assignments and timetables suitable for dynamic operations and detailed simulation. This limitation means that, although ExMAS can effectively generate high-quality potential ride-pooling combinations, applying it to practical fleet operation simulation and multi-city policy evaluation requires an additional mechanism capable of converting these ride-pooling solutions into executable, conflict-free vehicle schedules. Without such a mechanism, it is difficult to accurately assess the operational efficiency, service level, and interaction with public transport of SAV systems under real road networks and dynamic demand, which are the core concerns of this study. Therefore, this research is dedicated to developing and integrating a supplementary mechanism aimed at converting the ride-pooling solutions generated by ExMAS into executable, conflict-free vehicle schedules, thereby supporting detailed simulation analysis and multi-city comparison in this study.

In summary, current research on SAV fleet management mainly relies on two categories of methods. The first category, represented by general-purpose traffic simulation platforms such as MATSim and FleetPy, is capable of detailed simulation of large-scale transportation systems, individual travel behavior, and the dynamic evolution of traffic flow, making it suitable for evaluating the overall operation of SAVs in complex urban environments. The second category centers on mathematical optimization methods, including mixed-integer linear programming, mixed-integer nonlinear programming , graph theory models, and compartmental models. These methods can theoretically provide precise solutions for fleet size, scheduling, and routing optimization, and are suitable for flexible modeling and global optimization of operational parameters, constraints, and system objectives.

Table 2.3: Summary of Methods and Limitations

| **Article** | **Method** | **Limitations** |
|---|---|---|
| (Jin et al., 2021) | Mathematical modeling | Delays in matching and fleet size limit operation. Requires data calibration for critical densities in the fundamental diagram. |
| (Kucharski and Cats, 2020) | Mathematical modeling | The model assumes static demand and supply, ignores real-time traffic dynamics and demand elasticity, and does not consider fleet dispatching, rebalancing costs, or operational factors such as deadheading and driver incentives. |
| (Qu et al., 2022) | Mathematical modeling | Does not consider passenger satisfaction, vehicle charging and maintenance, supply-demand imbalance, repositioning of SAVs. |
| (Wang et al., 2018) | Simulation | The impact of different vehicle capacities was not considered, nor were passenger preferences for ride-sharing. |
| (Seo and Asakura, 2022) | Mathematical modeling | The model cannot separately identify individual traveler path flows and routes; demand is unknown or uncertain, potentially overestimating SAVs efficiency. |
| (Monteiro et al., 2021) | Simulation + Mathematical modeling | Factors such as weather and traffic incidents, which could impact vehicle availability, are not considered in the simulation. |
| (Fan et al., 2023) | Mathematical modeling | Assumes that travel demand is known and fixed, does not incorporate passenger waiting time, considers only non-ride-pooling services, and simplifies the traffic environment. The model does not reflect the dynamic nature of real-world demand, ride-pooling mechanisms, or the complexity of multimodal transportation. |
| (Balac et al., 2020) | Simulation + Mathematical modeling | Considers only car trips within the city of Zurich as the source of demand, assumes static and a priori known demand, does not analyze mode shift or the formation of new demand, and does not consider external inflows or unpredictable travel events, which limits the applicability and realism of the conclusions. |

Table 2.3: Summary of Methods and Limitations

| Article | Method | Limitations |
|---|---|---|
| (Militão and Tirachini, 2021) | Simulation | The study relies on simulation data from Munich and specific vehicle assignment strategies. The generalizability of the conclusions and their applicability to more complex scheduling algorithms require further validation. |
| (Vosooghi et al., 2019) | Simulation | Does not consider dynamic pricing strategies, the vehicle rebalancing algorithm requires further optimization, and the charging needs of electric SAVs and the layout of charging stations are not included in the simulation, which affects the practical applicability of the results. |
| (Boesch et al., 2016) | Simulation | Assumes static travel demand and idealized traffic conditions, does not consider ride-pooling, vehicle relocation, energy constraints, or demand hotspot optimization, and the conclusions are based only on the Zurich area, thus limiting regional applicability. |

Table 2.3 summarizes the main methodologies and limitations of the relevant literature. It is evident that, in current research on SAV operational strategies, mathematical optimization models (such as (Jin et al., 2021; Kucharski and Cats, 2020; Qu et al., 2022; Seo and Asakura, 2022; Fan et al., 2023)) have advantages in precisely solving for the optimal solution under specific objectives. However, when facing large-scale, dynamically changing urban transportation systems and the complex spatiotemporal constraints and multi-agent interactions in SAV operations, these models often encounter excessive computational complexity, difficulty in capturing all operational details, or excessive simplification of real-world scenarios.

In contrast, simulation platforms (such as (Wang et al., 2018; Monteiro et al., 2021; Balac et al., 2020; Militão and Tirachini, 2021; Vosooghi et al., 2019; Boesch et al., 2016)) offer high scalability and flexibility, and can effectively reproduce the dynamic evolution of complex transportation systems. However, in terms of generating ride-pooling solutions, mainstream simulation platforms mostly rely on heuristic or rule-based matching methods, making it difficult to systematically enumerate and evaluate all highly attractive ride-pooling combinations, and also challenging to achieve theoretical global optimality. The results are relatively sensitive to specific simulation settings and parameter rules, and there are certain limitations in terms of explanatory power and parameter controllability.

## 2.3 Research Gaps

Based on the preceding literature review, the following points summarize three gaps in the existing literature that motivate the research focus of this study:

- **Lack of Multi-City Heterogeneity Analysis:** Most studies use a single city as a case study (for example, Vosooghi et al. (2019) focuses on Rouen, France), or represent city differences using simplified indicators, and thus fail to fully reveal the combined impact of urban structure and demand distribution heterogeneity on SAV operational efficiency. For instance, Vosooghi et al. (2019) evaluates system performance by comparing preset fleet sizes, but lacks a systematic analysis of how changes in city characteristics affect operational strategies. This limitation restricts the applicability of research findings in diverse urban contexts, especially since SAV ride-pooling efficiency and fleet management strategies may vary significantly across cities of different scales and forms.

- **Insufficient Modeling of Multimodal Integration:** Although some literature has examined the impact of SAVs replacing traditional private cars on the transportation system (Fan et al., 2023; Jin et al., 2021), there is relatively little research on the competitive relationship between SAVs and public transport. This limits the understanding of how SAVs can achieve overall transportation system balance in a multimodal environment.
- **Gap Between Potential Matches and Executable Schedules:** A significant gap persists between theoretical ride-pooling algorithms that identify potential matches and the practical, conflict-free dispatching of vehicles(Kucharski and Cats, 2020). This disconnect leads to a systematic overestimation of achievable ride-pooling rates and a corresponding underestimation of the required fleet size, as theoretical potential fails to account for the spatio-temporal constraints of vehicle availability and travel time between tasks.

Consequently, several other important research areas, while identified as gaps in the literature (see Table 2.3), are considered beyond the scope of this work. For instance, this study does not delve into the modeling of dynamic demand, an area where many studies assume static conditions (e.g., (Kucharski and Cats, 2020; Boesch et al., 2016)), nor does it incorporate the impact of real-time disruptions like traffic incidents (Monteiro et al., 2021). Furthermore, specific operational complexities such as the optimization of vehicle rebalancing algorithms (Vosooghi et al., 2019) and the charging strategies for electric SAV fleets (Qu et al., 2022) are not addressed.

# Chapter 3

# Methodology

This Chapter aims to present a comprehensive simulation framework developed for analyzing SAV fleet operations. This framework, illustrated in Figure 3.1, meticulously models the sequential operational stages: beginning with initial travel demand and ExMAS-based ride option generation, proceeding through traveler mode choice that incorporates public transport alternatives, and culminating in an optimized vehicle assignment achieved via a two-stage process integrating efficient pre-filtering with global optimization. The elaboration of this framework begins with the travel demand generation model, which forms the foundation and input for subsequent analyses. Following this, a detailed exploration of the core mechanisms of SAV fleet operations will be undertaken. This involves several key steps: initially, potential ride options are generated from the travel demand using the ExMAS method. Subsequently, travelers' mode choice decisions are modeled, considering public transportation alternatives, to determine the specific trips allocated to the SAV system. For these SAV trips, a two-stage vehicle assignment process is then applied, where the first stage employs event-based vehicle matching to establish a decision space of feasible vehicle-to-ride pairings, and the second stage utilizes a global optimization to select the optimal assignments, thereby determining the final vehicle timelines and chosen rides. These interconnected stages—spanning demand generation, ride option creation, mode choice modeling, and culminating in the optimized assignment of vehicles—collectively form the comprehensive methodology presented in this section. This integrated approach is designed to facilitate a thorough analysis of SAV fleet operations and their potential to enhance overall urban transportation system efficiency and service levels.

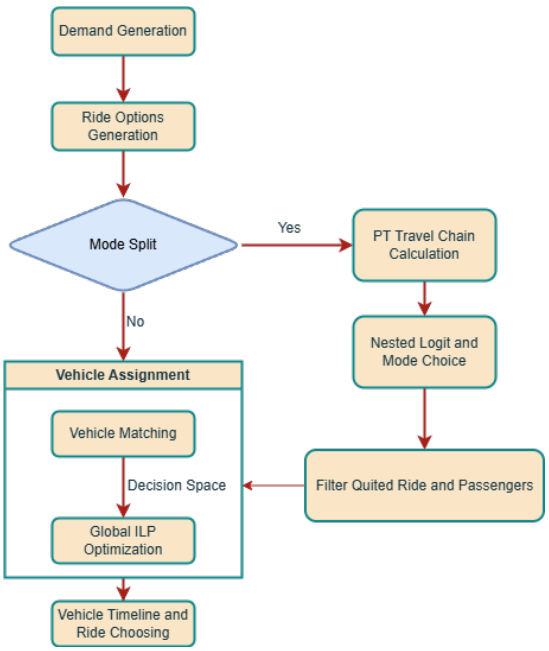



Figure 3.1: simulation framework

## 3.1 Demand Generation

This study adopts a case study approach to evaluate the optimal deployment strategies of shared automated vehicles in different city scales. Specifically, this study focuses on cities in the Netherlands. The study includes 37 cities of varying sizes, ranging from small cities (e.g., Almere) to large cities (e.g., Amsterdam), classified by population. The selection of cities is based on the ODiN dataset, which provides detailed travel data. Only cities with more than 1,000 travel records are selected to ensure data reliability and representativeness.

The demand utilized in ExMAS is based on the simulated travel data for Amsterdam from the albatross datasetArentze and Timmermans (2004). However, since this study covers multiple cities, it is necessary to generate full-day demand profiles that fit the characteristics of each city. After determining the total number of trips to be simulated for a city, the demand generation process is divided into two main components: calibration of travel characteristics and the generation of spatiotemporal information for trip origins and destinations.

### 3.1.1 Day Demand Scale

The travel demand in this study is based on the travel survey data from the ODiN dataset (voor de Statistiek , CBS) in the Netherlands, considering the travel demand within a 24-hour period. The calculation method for travel demand in this study is as follows:

$$D = Pop \cdot f \cdot s \tag{3.1}$$

where $D$ is the travel demand, $Pop$ is the population, $f$ is the average driving frequency, and $s$ is a scale factor. For example, the population of Amsterdam is 931,000, and the average driving frequency is 0.79 times per person per day (data from Centraal Bureau voor de Statistiek (CBS)). Therefore, the estimated daily travel demand is 6,705,000 trips. By adapting a downscaling method, with 1% of total estimated demand is considered, reducing it to 6,705 trips. To ensure comparability between different cities, the same scaling method is applied to small and medium-sized cities, resulting in a demand range of 1,000 to 4,000 trips in small and medium-sized cities.

It is important to clarify the meaning of this 1% demand. It is not a share of any specific mode's market penetration (e.g., taxis or SAVs). It is, by definition, exactly 1%of all urban travel behavior—including all trip purposes, all spatial and temporal distributions, and all original modes in the real world.

However, once this 1%is extracted, the modeling framework assumes a fully hypothetical environment in which only SAVs are available. That is, the entire simulation operates under the assumption of 100% SAV penetration, meaning all trips within the simulated sample are carried out using SAVs. This does not reflect current modal distributions in real cities; rather, it is a modeling boundary set to study the behavior and performance of a full SAV system in isolation.

Therefore, the simulated demand does represent 1%of actual urban travel behavior, but within a fictitious system in which every traveler uses SAVs and no other travel modes exist or influence the network. This modeling choice enables a focused analysis of SAV system operations without the interference of mixed-mode dynamics.

Nevertheless, it is necessary to mention that in studies where real-time traffic dynamics and congestion feedback are simulated—such as agent-based models using MATSim—downscaling the population (e.g., to 1%) may lead to distorted outcomes.Llorca and Moeckel (2019) showed that small-scale simulations often yield higher average travel times, unrealistic routing patterns, and unstable convergence behavior. These limitations arise because traffic dynamics are highly sensitive to density, vehicle interactions, and local congestion thresholds. However, this study does not simulate such dynamics. The allocation is static, offline, and non-congested. No network feedback or time-dependent capacity constraints are included. Therefore, the traffic-related biases associated with downscaling in dynamic simulations are not applicable here.

Furthermore, to evaluate whether using only 1% of total demand may introduce distortion in the simulation results, this study includes a dedicated sensitivity analysis of demand scale in small cities, $s$ was systematically increased from 1% to 10% , see details in Section 5.3.3.

### 3.1.2 Generation of Spatiotemporal Information for Origins and Destinations

To ensure that the simulated demand accurately reflects the real-world travel patterns of each city, the key statistical distributions used in this generation process, (hourly trip proportions and trip distance distributions for both commuting and non-commuting trips), are calibrated using the ODiN dataset(voor de Statistiek , CBS). The detailed methodology for this calibration is provided in Appendix G.2. This ensures that the generated demand realistically reflects empirical travel patterns.

The road network is extracted from OpenStreetMap (OSM)(OpenStreetMap, 2024), and all network nodes are considered as potential origins and destinations for trips. To better represent commuting patterns, additional area attributes are also obtained from OSM, allowing all nodes to be classified as work zones, residential zones, or other types(see Appendix G.1).

The demand generation process allocates trips spatially for each hour, following a structured methodology. For each trip, the origin is selected based on its type and the time of day. Commuting trip origins during the morning peak (6:00–9:00) are randomly selected from residential zones, while evening peak (16:00–19:00) origins are selected from work zones. For commuting trips outside these peak periods, which lack a single dominant direction, the model first probabilistically determines the trip's orientation. Each off-peak commute is randomly assigned as either a residential-to-work or a work-to-residential journey, each with a 50% probability. Once the direction is established, origins and destinations are selected from the respective zones. This method acknowledges the bi-directional nature of off-peak commutes while maintaining the structural integrity of work-related travel. For non-commuting trips, origins are selected randomly from any node in the network.

Once an origin is set, the destination is determined. First, a target travel distance is probabilistically selected from the corresponding empirical distance distribution (commuting or non-commuting)(see AppendixG.2). The system then assembles a pool of candidate nodes that are located at this target distance from the origin. This pool is then filtered based on zone type: for a morning commute, only nodes in work zones are kept; for an evening commute, only nodes in residential zones are kept. A final destination is then chosen randomly from this filtered set of feasible candidates.

To ensure trip generation is successful for every request, a fallback procedure is implemented. If no suitable destination is found for the initial target distance, the system iteratively expands the search to adjacent distance bins. If this process also fails, a destination is selected randomly from the entire network as a final resort. The framework also supports confining demand generation to specific hours for targeted analyses, and a maximum trip distance of 15 km is enforced on all trips to exclude unrealistic outliers.

## 3.2 The Joint Cost Optimization Formulation

For convenience, all parameters and their definitions are detailed in Appendix B. The objective of the final optimization stage in this study is to select a set of ride-vehicle assignments that satisfies all passenger travel demands while minimizing the system's total joint cost. This problem is formulated as the following integer linear program:

$$\min \quad \sum_{(r,v)\in \text{Pairs}} c_{r,v} x_{r,v} \tag{3.2}$$

$$\text{s.t.} \quad \sum_{(r,v)\in \text{Pairs}_p} x_{r,v} = 1 \quad \forall p \in P \tag{3.3}$$

$$x_{r,v} \in \{0,1\} \quad \forall (r,v) \in \text{Pairs} \tag{3.4}$$

where $x_{r,v}$ is a binary decision variable that equals 1 if the pre-computed assignment of vehicle $v$ to ride $r$ is selected, and 0 otherwise. The set of all feasible ride-vehicle pairs generated in the preprocessing stage is denoted by Pairs. The core constraint 3.3 ensures that each passenger $p$ from the set of all passengers $P$ is served by exactly one ride-vehicle assignment, by summing over the subset of pairs $\text{Pairs}_p$ that include passenger $p$.

The cost of each pair $c_{r,v}$ combines the weighted operator cost and the total user disutility, and is specifically defined as:

$$c_{r,v} = \alpha \cdot (f_{bal} \cdot O_{r,v}) + U_r \tag{3.5}$$

Here, $U_r$ represents the total user disutility for all passengers in ride $r$ (the detailed calculation is provided below), and this component mainly depends on the characteristics of the ride itself, such as passenger composition, route, and timing. $O_{r,v}$ denotes the operator cost incurred by assigning vehicle $v$ to serve ride $r$ (see Section 3.2 for details), which is related to the specific vehicle performing the task (e.g., whether a new vehicle needs to be activated, vehicle waiting time, etc.). As shown in Equation (3.5), to ensure comparability between user disutility and operator cost, the operator cost $O_{r,v}$ is multiplied by an internal balance factor $f_{bal}$, calculated as $f_{bal} = (\max(U_{pax}) - \min(U_{pax}))/(\max(U_{veh}) - \min(U_{veh}))$. The inclusion of this factor is a crucial normalization step that aligns the different numerical scales of user disutility and operator cost, ensuring that the subsequent weighting by $\alpha$ reflects a true policy trade-off rather than a technical correction for scale imbalance. Subsequently, for the purposes of sensitivity analysis and to evaluate different operational strategy preferences, the adjusted operator cost is further multiplied by a manually set weight factor $\alpha$ (corresponding to $\alpha$ parameter in Table 4.3, with a default value of 1), allowing the relative importance of operator cost in the total cost to be tuned during simulation.

The following contents provide a detailed explanation of the calculation methods for user cost $U_r$ and operator cost $O_{r,v}$.

**Non-Shared Ride Disutility ($U_i^{ns}$)** Disutility arises primarily from the time costs associated with the trip:

$$U_i^{ns} = \underbrace{F_i^{ns}}_{\text{Fare}} + \underbrace{\beta_{ivt} t_i}_{\substack{\text{In-vehicle} \\ \text{Time Cost}}} + \underbrace{\beta_{wait} \Delta t_i^{wait}}_{\substack{\text{Initial Waiting} \\ \text{Time Cost}}} \tag{3.6}$$

where the non-shared fare is $F_i^{ns} = \pi \cdot l_i / 1000$.

**Shared Ride Disutility ($U_{i,r}^s$)** For shared rides, disutility includes the fare, time costs associated with the service process (in-vehicle travel, pickup delay, boarding/alighting) valued at the in-vehicle rate and adjusted for sharing willingness, and the initial waiting time cost:

$$U_{i,r}^s = \underbrace{F_{i,r}^s}_{\text{Fare}} + \underbrace{\beta_{ivt} \omega (\hat{t}_{i,r} + |\Delta t_{i,r}^p| + \Delta t^{ba})}_{\substack{\text{Service Process} \\ \text{Time Cost}}} + \underbrace{\beta_{wait} \Delta t_i^{wait}}_{\substack{\text{Initial Waiting} \\ \text{Time Cost}}} \tag{3.7}$$

where the shared fare is $F_{i,r}^s = \pi(1-\delta) l_i / 1000$.

The parameters are defined as follows. The base fare rate per kilometer is denoted by $\pi$, and $l_i$ represents the travel distance for trip $i$ in meters. The discount applied for shared rides is captured by $\delta$. The value of time associated with the service process, including in-vehicle time and any delays during the ride, is denoted by $\beta_{ivt}$ for passenger $i$, while $\beta_{wait}$ reflects the value of initial waiting time before boarding. The variable $t_i$ corresponds to the direct (non-shared) in-vehicle travel time, and $\Delta t_i^{wait}$ indicates the initial waiting time before the ride begins. The factor $\omega$ represents the passenger's willingness to share and is applied to the time-related components of the service process. In a shared ride $r$, $\hat{t}_{i,r}$ denotes the in-vehicle travel time for passenger $i$, and $\Delta t_{i,r}^p$ refers to the pickup delay caused by sharing. Finally, $\Delta t^{ba}$ is the fixed amount of time allocated for each passenger's boarding and alighting.

**Operator Cost** The operator cost $O_{r,v}$ reflects the operational expenses incurred by vehicle $v$ to provide ride $r$:

$$O_{r,v} = \underbrace{c_w t_{r,v}^w}_{\substack{\text{Vehicle} \\ \text{Waiting Cost}}} + \underbrace{(c_f \cdot \mathbb{I}(\text{new}_v))}_{\substack{\text{Fixed Vehicle} \\ \text{Activation Cost}}} + \underbrace{c_t t_{r,v}}_{\substack{\text{Driving} \\ \text{Cost}}} \tag{3.8}$$

Here, $c_w$ denotes the cost per unit time of vehicle waiting due to early arrival, and $t_{r,v}^w$ represents the total waiting time of vehicle $v$ before starting ride $r$. The term $c_f$ refers to the fixed cost incurred when activating an inactive vehicle, while the indicator function $\mathbb{I}(\text{new}_v)$ equals 1 if vehicle $v$ is newly introduced to the system for this ride, and 0 otherwise. The driving cost per unit time or per kilometer is denoted by $c_t$, and $t_{r,v}$ stands for the total duration or driving time for vehicle $v$ to complete ride $r$, including the travel time to the pickup location.

This comprehensive cost function $c_{r,v}$ allows the optimization framework to balance operational efficiency and user-perceived costs (including time, delay, and fare) when determining the optimal set of ride-vehicle pairs and the required fleet size.

When the system determines that no existing idle vehicle is capable of serving a ride $r$ within the passenger's maximum acceptable delay, a new vehicle $v$, which was previously in an inactive state, is activated. In such instances, the cost is calculated as:

- **Fixed Activation Cost:** The operator cost $O_{r,v}$ includes a fixed activation cost for new vehicles, represented by the term $c_f \cdot \mathbb{I}(\text{new}_v)$, where the indicator function $\mathbb{I}(\text{new}_v)$ equals 1 if a new vehicle is activated. The value $c_f$ corresponds to the parameter New Vehicle Fixed Cost.
- **Initial Waiting and Driving Cost:** The system uses the preset parameter Average Waiting Time for New Vehicle (denoted as $t_{wait}^{new}$) to simulate the time required for a newly activated vehicle to reach the passenger's initial location, which is set based on the average pick-up time across all cities. This time affects cost calculations in two ways:
  1. It constitutes the initial waiting time for passenger $i$ in the ride, $\Delta t_i^{wait} = t_{wait}^{new}$, and is included in the total user disutility $U_r$ via the term $\beta_{wait}\Delta t_i^{wait}$ (see Equations 3.6 and 3.7).
  2. The same $t_{wait}^{new}$ is also treated as the driving time for the newly activated vehicle $v$ to reach the first passenger's pick-up point (i.e., the pick-up time $t_{pick}$), and is thus included in the operator's driving cost term $c_t t_{r,v}$.
- **Vehicle Waiting Cost:** For newly activated vehicles, the vehicle waiting time upon arrival at the first passenger's pick-up point, $t_{r,v}^{w}$, is assumed to be zero, as these vehicles are dispatched on demand and do not need to wait after reaching the passenger's location.

It is important to emphasize that this approach to calculating new vehicle costs relies on two fixed input parameters (*New Vehicle Fixed Cost* and *Average Waiting Time for New Vehicle*). As discussed later in the results analysis (see, for example, the discussion of $p_{init}$ in Section 5.1), excessive reliance on new vehicle activation may introduce simulation bias due to the simplified treatment of these fixed parameters. Therefore, in model optimization and parameter calibration, it is preferable to provide sufficient vehicle resources to avoid the need for new vehicle activation, which is an important consideration for improving the realism of simulation results.

## 3.3 A Two-Stage Algorithm for Solving the Optimization Problem

As mentioned in Section 1, because the scope of this research focuses on SAV fleet management strategies based on predetermined demand, ExMAS, as an offline algorithm, offers a more rational and efficient ride-sharing optimization solution compared to other simulation methods with on-demand service scenarios Kucharski and Cats (2020). However, the standard ExMAS framework estimates the optimal fleet size based on the peak demand observed in terms of concurrent trips, without explicitly incorporating vehicle dispatching and assignment in complex urban environments. This approach may deviate from real-world, all-day operational scenarios. To more accurately simulate vehicle scheduling and determine the actual required fleet size, this study has adapted the ExMAS framework. The core innovation lies in a two-stage ride-vehicle matching algorithm: the first stage employs an efficient filtering mechanism to reduce the search space of candidate ride-vehicle assignments for the second stage, and the second stage then applies a global optimization model to select the final, conflict-free plan from this set.

### Generation and Filtering of Potential Shared Rides

The first stage focuses on generating all feasible shared ride combinations based on individual travel demands, including origin, destination, departure time,and trip distance. This process primarily uses the original ExMAS algorithm, which systematically constructs candidate ride-pooling groups. A shared ride is considered attractive to passengers only if the cost reduction from fare discounts is sufficient to overcome the disutility caused by additional time costs caused by sharing the ride, such as detours or delays. The construction of shared rides starts with pair-wise combinations and incrementally expands to larger groups, up to the maximum vehicle capacity(see Appendix C). Notably, single-passenger direct rides are also treated as a special case within this framework.

If the model is configured to account for competition with public transport, a further filtering step is introduced. At this stage, the system applies a nested logit model to estimate the probability that each

passenger in a candidate ride would opt for public transport instead of the SAV service. Mode choice for each passenger is then determined through probabilistic sampling, as will be detailed in subsequent sections3.4. If any passenger within a candidate ride is found to prefer public transport, that ride is marked as infeasible for SAV system and removed from the candidate rides set. The system also records all passengers associated with such removed rides. After evaluating all candidate rides, if a passenger consistently prefers public transport across all their potential ride options, that passenger is excluded from the demand pool for SAV services. This filtering process is designed to eliminate passengers and ride combinations that are unlikely to select SAV, thereby streamlining the subsequent vehicle assignment and optimization steps.

## Vehicle Entity Modeling

To accurately simulate fleet operations and evaluate system efficiency, the model explicitly represents each service vehicle as an individual entity. These vehicle entities constitute the core supply side of the ride-pooling service.

- **Vehicle Initialization:** At the beginning of the simulation, the system initializes a fleet with a predetermined number of vehicles, which are randomly distributed across nodes on the map and are available for service from the start of the simulation.
- **Vehicle States:** Throughout the simulation, each vehicle is always in one of several mutually exclusive operational states:
  - **Idle:** The vehicle is not currently assigned to any passenger task, is available, and is waiting for new assignment instructions.
  - **Busy:** The vehicle is engaged in one or a sequence of consecutive passenger tasks, including traveling to pick-up points and transporting passengers. In this state, the vehicle cannot be reassigned until all current tasks are completed.
  - **Inactive:** The vehicle has not yet been deployed and remains in reserve. Only when the active fleet cannot meet demand will inactive vehicles be activated and switched to the busy state.

  Vehicle states are dynamically updated according to task assignments and completions. For example, an idle vehicle becomes busy upon accepting a new task; after completing its tasks, if there is no immediate subsequent assignment, it returns to the idle state.
- **Vehicle Activity Timeline Tracking:** To accurately determine the availability and location of each vehicle at any given time, the system maintains a detailed activity timeline for every vehicle. This timeline records the full sequence of assigned tasks, along with the precise start and end times for each. The timeline is dynamically updated as vehicles complete tasks and potentially accept new ones, with the expected future availability of each vehicle adjusted accordingly. Tracking these activity timelines is essential for making effective assignment decisions and avoiding scheduling conflicts, especially when a vehicle appears idle but is already reserved for future tasks. By analyzing the activity timeline, key performance indicators such as total active time and service time for each vehicle can be calculated.

## Stage 1: Vehicle Assignment and Ride Option Generation

This stage is central to the system, aiming to identify a high-quality and feasible set of ride-vehicle assignment pair candidates for subsequent global optimization. The core objective is to quantify the impact of different assignment schemes on overall total cost while ensuring that all passenger service quality constraints(all passengers need to be served once and only once) are satisfied. This approach significantly reduces the search space for the final optimization. The detailed procedure is as follows:

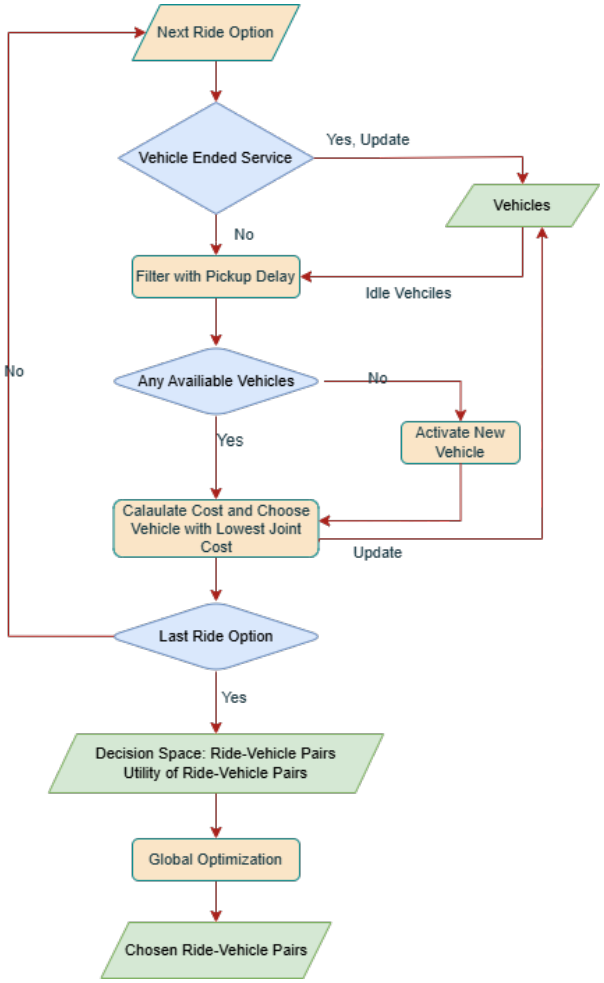



Figure 3.2: vehicle assignment flowchart

1. **Temporal Processing of Rides:** All potential shared rides are sorted according to the earliest requested departure time of their primary passenger. The algorithm processes each ride sequentially in time, simulating the dynamic evolution of the system. This approach ensures that vehicle status is evaluated based on a basic spatiotemporal law, which is fundamental for determining whether a vehicle can serve a new request at a specific moment.

2. **Dynamic Vehicle State Update:** Before evaluating each ride, the system checks the status of all busy vehicles. If a vehicle has completed its previous task (i.e., the end time is earlier than or equal to the current ride's request time), its status is updated to idle, its location is set to the endpoint of the last task, and its availability is recorded. This guarantees that assignment decisions are always based on the most up-to-date information.

3. **Evaluating Feasibility and Cost of Idle Vehicles:** To control fleet size, the core principle of the algorithm is to prioritize the use of currently idle vehicles. For each ride under consideration, the system iterates through all idle vehicles, first calculating the expected travel time from the vehicle's current location to the pick-up point of the primary passenger. Based on this, the expected arrival time is estimated and compared with the passenger's requested departure time, yielding the potential passenger departure delay and vehicle waiting time. A feasibility check is then performed: only if the calculated passenger departure delay is less than the passenger's maximum acceptable threshold is the idle vehicle considered a feasible candidate. For all idle vehicles passing this check, the system further computes the incremental joint cost of assigning the vehicle to the ride. This increment is a composite metric quantifying the marginal impact of the assignment on both the passenger and the system, including the passenger's time cost due to departure delay (valued according to their value of time), the additional travel cost for the vehicle, and the waiting cost if the vehicle for arriving early (weighted by a waiting cost penalty factor). All these cost components are balanced by appropriate factors to ensure a unified evaluation standard(see Section5.1.4). This

mechanism of prioritizing idle vehicles and filtering based on strict time constraints (maximum delay) not only quickly eliminates options that do not meet service quality requirements and reduces the search space, but also forms the basis for implicit vehicle repositioning: as long as the time constraint is satisfied, even idle vehicles located farther away may be dispatched, thus maximizing the utilization of the existing fleet.

It is important to note the core strategy adopted in this study to avoid bias in fleet size estimation. The model always attempts to use feasible idle vehicles to serve ride requests first, and only activates new vehicles when no idle vehicle can reach the passenger within the maximum acceptable delay. By strictly prioritizing feasible existing vehicles and incorporating all relevant costs (passenger delay, vehicle travel, vehicle waiting, and new vehicle activation) into the final global optimization objective (see Equation 3.2), the model can inherently seek solutions that minimize the actual number of deployed vehicles while meeting all demand. Allowing the use of new vehicles ensures that the system can always assign a vehicle to every ride, which, although rarely needed, is essential for system robustness in extreme cases.

4. **Activating New Vehicles as Backup:** Only when all idle vehicles have been considered and none can serve the ride within the passenger's maximum acceptable delay does the system activate a new vehicle to ensure service continuity. In this case, the system simulates the deployment of an new vehicle(inactive status). The passenger must accept a preset average waiting time for the new vehicle, which is counted as both passenger delay and vehicle pick-up time. The incremental joint disutility for this option is also calculated, including not only the passenger's waiting cost but also the fixed activation cost representing the investment in new resources. This process explicitly quantifies the marginal cost of deploying new resources as a fixed value, enabling the subsequent optimization to balance service efficiency and operational cost, and also simulates the real-world scenario of dynamically supplementing capacity as needed. Once a new vehicle is activated, its status is immediately updated to busy, and its activity timeline begins.

5. **Candidate Ranking and Matching:** Theoretically, it is possible to construct an exact optimization model that includes all potential ride-vehicle combinations. However, for real-world city-scale applications with large numbers of ride requests and vehicles, such a model would become computationally infeasible due to the exponential growth in the number of decision variables and constraints (especially those ensuring vehicle time continuity and avoiding spatiotemporal conflicts). To balance computational feasibility and solution quality, the model adopts a myopic assignment strategy that makes the best immediate assignment for each ride as it is processed sequentially, without considering the impact on future events. For each ride under consideration, all feasible vehicle(idle vehicles that can meet the passenger's maximum pickup delay constraint) assignment options are ranked according to the joint cost defined in Equation (3.5). The vehicle with lowest cost as the execution plan for that ride is then be chosen. Once a vehicle is assigned, its status is temporarily updated to busy, its expected availability is adjusted to the end of the ride, and the ride is recorded in its virtual activity timeline. All other feasible vehicles are immediately released for subsequent assignments.

   This approach of making optimal choices for each ride sequentially and assigning vehicles instantaneously approximates an event-based simulation process. While this method can estimate the minimum fleet size required to meet all passenger needs, its inherent nature, which is making assignment decisions without considering the potential for more optimal future ride sequences, means it may not always achieve the global optimum for such offline scenarioS. Nevertheless, this approach effectively mimics the real-world context where ride requests are made in real time and the platform must respond promptly.

In summary, this stage integrates temporal simulation, service constraint filtering, joint disutility evaluation, dynamic resource management, and event-driven vehicle assignment to enable a comprehensive simulation of the full-day ride-pooling system.

## Stage 2: Global Optimization and Assignment

**Integer Linear Programming Optimization:** As mentioned in Section 3.2, the core objective of the model is to minimize the joint cost function defined in Equation (3.5). The fundamental assumption behind this design is that a sustainable SAV service must strike a balance between operator efficiency and user attractiveness. Rather than simulating the complex decision-making of individual operators

and users, the model abstracts this trade-off into a cost-minimization problem. This approach makes it possible to systematically explore system performance under different parameter settings.

The final optimization stage is designed to select a globally optimal and consistent set of assignments from the pool of candidates generated previously. The first stage, as already detailed, produces a large number of candidate (ride, vehicle) pairs. The central challenge is that these candidate rides overlap, as a single passenger can be included in multiple, mutually exclusive ride options, such as a solo ride and several different shared rides.

This structure means the problem is inherently a set partitioning problem. This mathematical framework is ideally suited for selecting a collection of items that perfectly covers a required set of elements without any redundancy. In this context, the complete set of passengers ($P$) constitutes the "universal set" that must be partitioned, and the candidate rides, each containing a specific group of passengers, function as the available "subsets" for creating this partition. The binary decision variable $x_{r,v}$ represents the selection of a ride-vehicle pair. Constraint (3.3) then mathematically enforces the partitioning by mandating that for each passenger $p$, the sum of selected rides covering that passenger must equal one. This ensures the final plan is conflict-free and serves all travel demand.

The optimization problem is implemented in Python using the *PuLP* modeling library(Mitchell et al., 2011). The model is solved using the open-source Coin-OR Branch-and-Cut (*CBC*) solver, which is well-suited for large-scale integer programming problems(Forrest et al., 2024).

## 3.4 Integration with Public Transport

As emphasized by Kucharski and Cats (2020), the ExMAS algorithm can also be utilized to analyze the competitive or complementary relationship betwee shared mobility service and public transport. In this section, a Nested Logit model and a binary logit model are employed to allocate passenger choices between SAVs and public transport, thereby evaluating the competition structure between these two modes. The analysis in this study focuses specifically on all urban rail transit systems, given their extensive network coverage and comparable average speeds between different rail transport services.

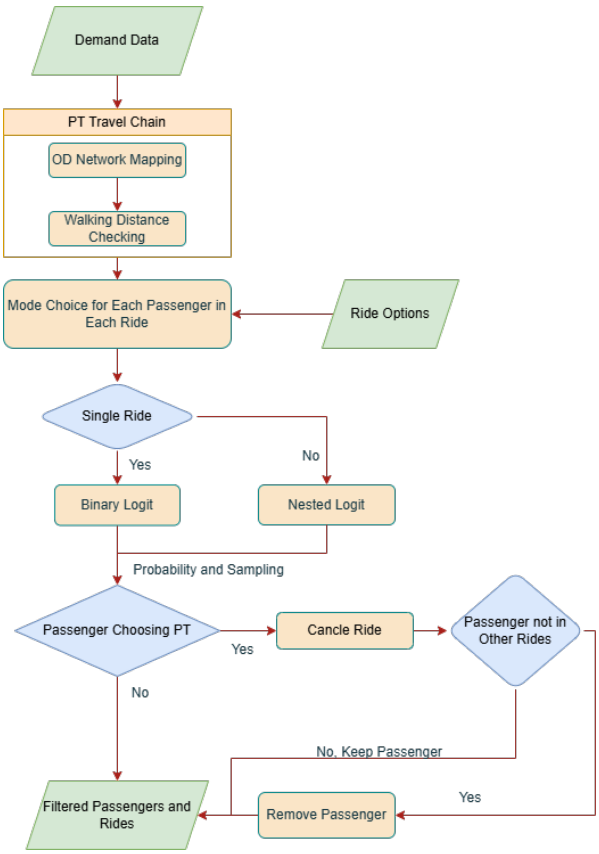



Figure 3.3: PT integration flowchart

### 3.4.1 Public Transport Cost Function Formulation

The construction of the cost function for public transport travel closely mirrors that of SAV, as it comprehensively incorporates both monetary and temporal costs. Specifically, the cost of public transport is formulated as follows:

$$U_{PT} = (1 - \delta_{PT}) \cdot (\text{Fare}_{fixed} + \text{Fare}_{dist} \cdot d_{PT}) + \beta_{PT,ivt} \cdot t_{PT} + \beta_{walk} \cdot t_{walk}$$

In this expression, $\delta_{PT}$ denotes the discount rate applied to the total fare of public transport, which is set to zero by default. The terms $\text{Fare}_{\text{fixed}}$ and $\text{Fare}_{\text{dist}}$ represent the fixed fare and the distance-based fare component, respectively, while $d_{PT}$ corresponds to the mainline travel distance of PT. The variable $t_{PT}$ captures the in-vehicle travel time, and $t_{\text{walk}}$ refers to the total walking time required for access and egress; in scenarios where public transport is unavailable, this term reflects the walking time from the origin to the destination. The parameters $\beta_{PT,ivt}$ and $\beta_{walk}$ quantify the value of in-vehicle time and walking time, respectively, thereby encapsulating travelers' sensitivity to these temporal costs.

This cost function establishes a rigorous foundation for the subsequent mode choice modeling framework.

### 3.4.2 Modeling Public Transport Travel Chain

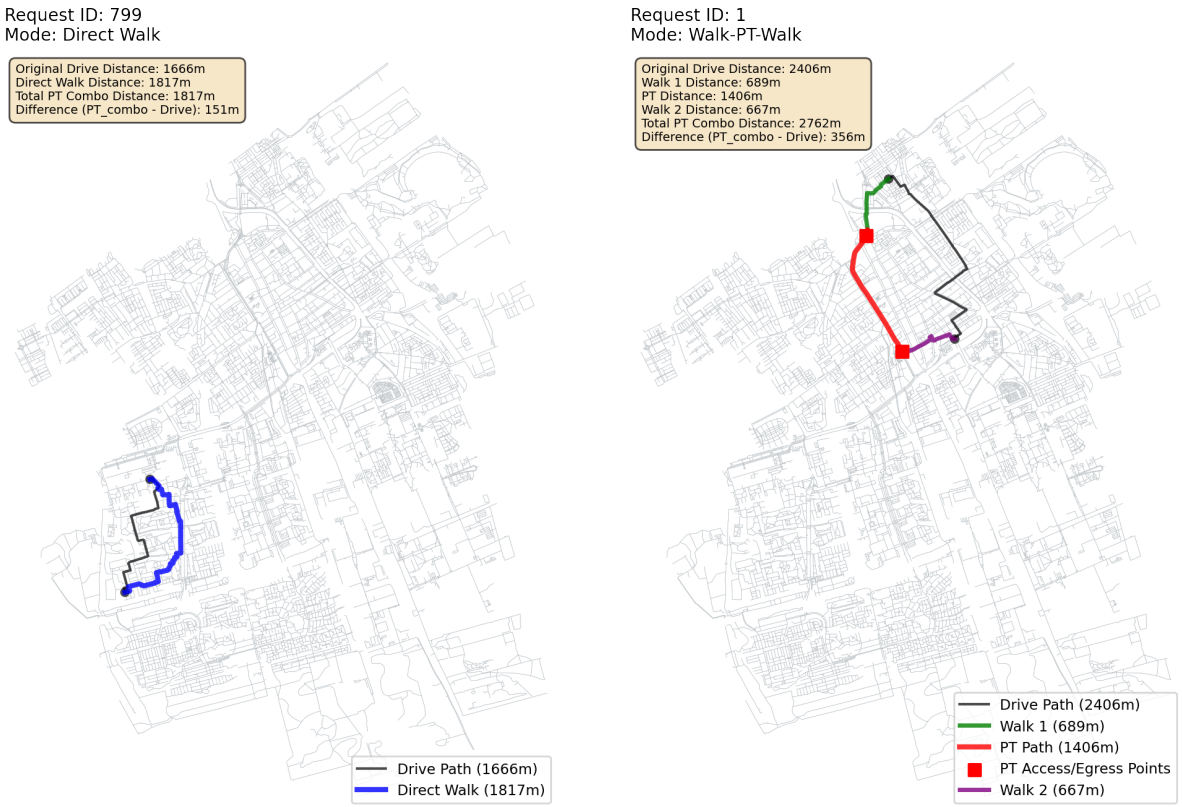
To accurately assess the service level and attractiveness of public transport, the model first constructs the complete travel chain for passengers using PT. This process relies on a detailed representation of the city's multimodal transport network. Specifically, the system integrates three key network layers: the road network serving SAVs, the public transport network comprising rail lines, and the walking network used to simulate passenger access and egress. All these networks are downloaded from OpenStreetMap data(see Appendix F).

For any given travel demand, the system maps it onto a potential public transport journey. This mapping process first identifies, based on Euclidean distance, the public transport stations closest to the trip's origin and destination, which are then designated as the starting and ending points of the public transport segment. Subsequently, the walking network is used to determine the access and egress paths connecting the trip origin to the public transport origin and destination to the trip destination. This involves locating the nearest nodes in the walking network to these four key locations and calculating the corresponding walking distances.

Based on the segmented journey distances obtained above (including both walking and PT mainline segments), and in conjunction with the average speed parameters of each mode, the system computes the total travel time for the potential PT journey. This total time is further divided into key components: in-vehicle PT time and total walking time, both of which serve as critical inputs for subsequent cost function calculations and mode choice analysis. As a simplification, PT waiting time is not considered at this model.

In addition, to prevent the model from considering unrealistic travel options, a filtering logic is developed to identify and convert infeasible PT routes into the pure walking mode.

This filtering logic is based on two core criteria. The first is the minimum PT segment distance threshold. If the calculated main PT segment distance is shorter than a threshold (set as 100 meters in this study), the system determines that the origin and destination are at the same stop or that the PT stops are mapped to the same stop on the opposite side of the road. In such cases, the PT segment is considered infeasible. The second criterion is the comparison of walking distances. The system compares the total walking distance of the PT combined route (i.e., the sum of access and egress walking) with the direct walking distance from origin to destination. If the former is significantly longer than the latter, detouring via public transport is deemed unreasonable. Figure 3.4 illustrates the segments of pure walking routes and PT modes.




(a) Example of a direct walking route

(b) Example of a public transport trip chain

Figure 3.4: Examples of direct walking route and public transport trip chain

It is important to note that the primary focus of this study is the competition between SAVs and public transport; therefore, a separate and complex cost function is not developed for the direct walking trips. Instead, walking is regarded as a basic filtering mechanism that excludes unreasonable short-distance SAV trips. The cost of walking is simplified to be directly determined by walking time, calculated as $U_{\text{walk}} = \text{VoT}_{\text{walk}} \cdot T_{\text{walk}}$, which is consistent with the cost calculation for the walking component in the public transport cost function.

At the final mode choice stage, for short-distance trips where direct walking is determined to be superior, the system calculates a selection probability based on the utilities of both SAV and direct walking, followed by a random draw(see Section 3.4.3). If the passenger chooses walking according to this probability, the corresponding trip is completely removed from the subsequent SAV matching pool and is not included in the public transport system. This approach effectively filters out short trips where SAVs lack a competitive advantage, allowing the subsequent matching optimization to focus on trips where SAVs provide genuine service value. As a result, the overall simulation outcomes are ensured to be both realistic and reasonable.

### 3.4.3 Mode Choice Model

This research employs two types of Logit models grounded in the theory of random utility maximization to simulate passengers' choices among shared SAV, non-shared SAV, and public transport.

For each potential passengers in shared ride options, the model calculates the utility values associated with three transportation modes. While for solo SAV ride options, the model calculates the non-shared SAV utility and PT utility only. The utility of shared SAV mode ($U_{\text{shared}}$) is computed by the ExMAS core algorithm, accounting for discounts, detours, delays, and other relevant factors. The non-shared SAV utility ($U_{\text{non-shared}}$) typically serves as the baseline utility, considering only the direct route's cost and time. The public transport utility ($U_{PT}$) is formulated as described in the previous section.

The model structure selection takes two different forms depending on the trip type. For solo SAV ride options, since the choice is limited to non-shared SAV and public transport, the standard Binary

Logit model is adopted. The probability of selecting mode $m$ is given by:

$$P(m) = \frac{e^{V_m}}{\sum_{k \in \{\text{non-shared}, PT\}} e^{V_k}}$$

where $V_m = -U_m$ denotes the utility of mode $m$, with $U_m$ representing the cost.

For shared SAV ride options, considering that shared and non-shared SAVs both belong to the SAV category and may exhibit higher similarity compared to PT, a nested Logit model is applied. This model structure is organized in two levels: in the upper level (Mode Choice), passengers first choose between the SAV nest and PT; in the lower level (SAV Service Choice), if the SAV nest is chosen, passengers then select between shared SAV service and non-shared SAV service.

The overall attractiveness of the SAV nest is determined by the Logsum of its constituent options: orrelation Analysis between City Characteristics and Operational Indicators

$$IV_{SAV} = \lambda_{SAV} \ln \left( e^{V_{\text{shared}}/\lambda_{SAV}} + e^{V_{\text{non-shared}}/\lambda_{SAV}} \right)$$

where $\lambda_{SAV}$ is the scale parameter for the SAV nest ($0 < \lambda_{SAV} \leq 1$), reflecting the correlation among alternatives within the nest. The upper-level choice probability is calculated analogously to the binary Logit:

$$P(\text{SAV Nest}) = \frac{e^{IV_{SAV}}}{\sum_{k \in \{\text{SAV Nest}, PT\}} e^{IV_k}} \quad \text{and} \quad P(PT) = \frac{e^{V_{PT}}}{\sum_{k \in \{\text{SAV Nest}, PT\}} e^{IV_k}}$$

(PT forms a single nest with a scale parameter of 1, so $IV_{PT} = V_{PT}$.)

The final probability of selecting a specific SAV mode is obtained by multiplying the conditional probability by the nest choice probability:

$$P(\text{shared}) = P(\text{SAV Nest}) \cdot \frac{e^{V_{\text{shared}}/\lambda_{SAV}}}{e^{V_{\text{shared}}/\lambda_{SAV}} + e^{V_{\text{non-shared}}/\lambda_{SAV}}}$$

$$P(\text{non-shared}) = P(\text{SAV Nest}) \cdot \frac{e^{V_{\text{non-shared}}/\lambda_{SAV}}}{e^{V_{\text{shared}}/\lambda_{SAV}} + e^{V_{\text{non-shared}}/\lambda_{SAV}}}$$

In the stochastic choice and filtering process, the system performs random draws based on the probability distributions calculated above for each passenger in potential shared ride options to determine their selected modes. When any passenger in a shared ride option chooses public transport, the system automatically marks the entire shared option as infeasible and removes it from the generated ride options (Section 3.3). Additionally, these passengers who select public transport are excluded from requests.

# Chapter 4

# Simulation Scenarios

This Chapter introduces experiment plan designed to evaluate the performance of the SAV system across different urban scales and to assess the competitive relationship between SAV and PT systems. This approach will be used to derive insights for both scenarios, thereby supporting the formulation of operational strategies.

The subsequent content will first introduce the key parameters and indicators, followed by a description of the simulation design to answer research questions.

## 4.1 Demand Generation and Data Source

To ensure the simulation is grounded in realistic urban conditions, this study synthetically generates travel demand for each of the 37 municipalities. The generation process is designed to replicate key real-world travel patterns and is based on a two-step methodology: functional zone classification and spatio-temporal characteristic calibration.

First, urban space is categorized into *Residential*, *Work*, and *Other* functional zones using land use data from OpenStreetMap (OSM). This classification allows for the simulation of typical commuting patterns, such as morning and evening peak travel flows between home and work locations. Subsequently, the temporal and spatial characteristics of travel are calibrated using the ODiN dataset from the Dutch statistical office (CBS). This involves adjusting the simulated demand to match observed data in terms of hourly trip distribution, the proportion of commuting trips per hour, and the distribution of trip distances.

This calibrated approach produces a solid demand data foundation for the subsequent simulation of the SAV system, ensuring that the analysis results accurately reflect the potential performance of SAVs in specific urban environments. The detailed methodology for zone classification, along with illustrative figures and a full discussion of the calibration results for the representative case of Amsterdam, is provided in Appendix G.

## 4.2 Key Model Parameters

The model incorporates several key parameters that influence system behavior. These are categorized into two groups for clarity. The first group, detailed in Table 4.1, consists of calibration parameters. These parameters, while fundamental to the model's operation, lack direct real-world data for their values. Therefore, they are assigned baseline values through a dedicated calibration process (detailed in Section 4.5.1) to ensure the model operates under reasonable and consistent assumptions.

Table 4.1: Model Calibration Parameters

| **Parameter Name** | **Symbol** | **Unit** | **Side** | **Description** |
|---|---|---|---|---|
| Initial Proportion | $p_{init}$ | - | Operator | The proportional factor used to calculate the initial fleet size, where the initial fleet size is determined by multiplying the total number of passengers by this factor. |
| Waiting Cost | $c_w$ | euros/s | Operator | The unit time cost of operator incurred by a vehicle waiting for passengers due to early arrival, used to penalize premature arrivals. |
| Utility Balance | $\alpha$ | - | Operator | An additional balancing factor used to adjust the weight of operator costs in the total utility calculation. This factor works in conjunction with the balance factor to adjust the weight of the operator's utility. |

The parameters detailed in Table 4.3 are specifically chosen for the sensitivity analysis because they directly correspond to the research sub-questions of this study. Systematically varying these parameters enables a targeted investigation of their individual impacts on system performance. It is important to note that the *Willingness to Share Resistance Factor* serves a dual role; it is first calibrated to establish a city-specific baseline reflecting realistic sharing preferences, and is subsequently varied in the sensitivity analysis to explore its broader impact on system outcomes.

Table 4.3: Sensitivity Analysis Parameters

| Parameter Name | Symbol | Unit | Side | Description |
|---|---|---|---|---|
| Maximum Pickup Delay | $\Delta t^{p,\max}$ | s | Operator | The maximum pickup delay time that passengers can tolerate. This parameter serves as a critical operator-defined service level parameter, acting as a hard constraint that directly limits the highest acceptable passenger waiting time and influences vehicle assignment feasibility. |
| Shared Discount | $\delta$ | % | Operator | The fare discount percentage for passengers when choosing shared rides. |
| Willingness-to-Share Resistance Factor (WtSR) | $\omega$ | - | User | A factor reflecting passengers' resistance for additional time costs (such as delays and detours) associated with shared rides, influencing the calculation of shared utility and the feasibility of ride-sharing. |
| Scale Factor | $s$ | - | User | The percentage of simulated demand relative to the total actual travel volume. |
| PT Discount | $\delta_{PT}$ | % | PT | The discount percentage applied to the total fare of public transport. |
| Average Speed | $v_{avg}$ | m/s | Operator | The average speed of SAVs, used to calculate vehicle travel times and various time-related KPIs. |

This study sets fixed values for parameters supported by research(see Appendix C), such as value of time and price, and parameters that can be validated through the ODiN dataset. These parameters will not be included in the analysis, as they serve as a basis for the system to accurately reflect reality.

## 4.3 Key Performance Indicators

To evaluate the performance of the proposed system, this study employs a series of key performance indicators. The following tables present the core KPIs that are most relevant to the research objectives and analysis outcomes. These indicators focus on fleet efficiency, system performance, ride-sharing effectiveness, and service quality metrics that directly address the research questions. Additional KPI definitions, including utility metrics and other secondary performance measures, are provided in Appendix E.

Table 4.5: Fleet Size Metrics

| KPI Name | Description |
|---|---|
| Total Vehicles | Number of unique vehicles used to serve the rides |
| Peak Concurrent Vehicles | Maximum number of concurrently serving vehicles |

Table 4.7: Efficiency Metrics

| KPI Name | Description | Formula |
|---|---|---|
| Vehicle Reuse Rate | Percentage of vehicles utilized for multiple trips | $\|V_{reused}\|/\|V_{used}\|$<br><br>Where $V_{reused}$ is the set of vehicles serving more than one ride. |
| Average Rides per Vehicle | Average number of trips handled by each vehicle | $\|R^*\|/\|V_{used}\|$ |
| Saved Miles Ratio | Proportion of total vehicle mileage saved compared to the shortest path mileage for passengers, modification on pax detour ratio from(Jin et al., 2021)<br>(How much mileage is saved by ride-sharing compared to the shortest route) | $(D_{direct} - D_{service})/D_{direct}$<br><br>$D_{service} = \sum_{(r,v)\in R^*} d_{service}(r)$,<br>$D_{direct} = \sum_{p\in Pax_{served}} d_{direct}(p)$. |
| Extra Mileage Ratio | Proportion of pickup mileage to total mileage (Jin et al., 2021)<br>(Additional distance traveled for pickup / total distance) | $D_{pick}/(D_{pick} + D_{service})$<br><br>$D_{pick} = \sum_{(r,v)\in R^*} d_{pick}$. |
| Time Utilization Rate | Ratio of vehicle service time to total active time | $T_{service}/T_{active}$<br><br>$T_{service} = \sum_{(r,v)\in R^*} t_{service}(r)$,<br>$T_{active} = \sum_{v\in V_{used}} (t_{end}^{last}(v) - t_{start}^{first}(v))$. |
| Moving Time Ratio | Ratio of vehicle moving time (service time + pickup time) to total active time | $(T_{service} + T_{pick})/T_{active}$<br><br>$T_{pick} = \sum_{(r,v)\in R^*} t_{pick}$. |

Table 4.9: Pooling Metrics

| KPI Name | Description | Formula |
|---|---|---|
| Pooling Ratio | Percentage of trips involving ride-sharing in the selected solution | $\|Pax_{shared}\|/\|Pax_{served}\|$<br><br>$Pax_{shared}$ is the set of passengers who actually joined a shared ride, and $Pax_{served}$ is the set of all served passengers. |
| Shared Passengers Ratio | Proportion of passengers who was arranged to shared rides from the original rides optionsKucharski and Cats (2020) | $\|Pax_{shared\_options}\|/\|Pax_{all}\|$<br><br>$Pax_{shared\_options}$ is the set of passengers who were offered shared ride options. |

Table 4.11: Waiting Time Metrics

| KPI Name | Description | Formula |
|---|---|---|
| Average Vehicle Waiting Time | Average waiting time for vehicles to pick up passengers per ride | $\sum_{(r,v)\in R^*} t_{wait\_veh}/\lvert R^*\rvert$<br>$t_{wait\_veh} = \max(0, t_{actual\_start} - t_{arrival})$. |
| Average Passenger Waiting Time | Average waiting time for passengers before being picked up | $\sum_{p\in Pax_{served}} t_{wait}(p)/\lvert Pax_{served}\rvert$<br>$t_{wait}(p) = \max(0, t_{pickup\_actual}(p) - t_{req}(p))$. |
| Average Pickup Time | Average travel time for vehicles to pick up passengers | $\sum_{(r,v)\in R^*} t_{pick}/\lvert R^*\rvert$<br>$t_{pick}$ is the travel time for vehicle v from its current position to the origin of ride r. |

Table 4.13: KPIs Used in PT Analysis

| KPI Name | Description | Formula |
|---|---|---|
| Average Quit PT Distance | Average PT segment distance for passengers switched to PT | $\sum_{p\in Pax_{PT}} d_{PT}(p)/\lvert Pax_{PT}\rvert$ |
| Average Quit Distance | Average original requested travel distance for passengers switched to PT | $\sum_{p\in Pax_{PT}} d_{origin}(p)/\lvert Pax_{PT}\rvert$ |
| Walk Ratio | Proportion of passengers switched to direct walking | $\lvert Pax_{walk}\rvert/\lvert Pax_{all}\rvert$ |
| PT Ratio | Proportion of passengers switched to PT | $\lvert Pax_{PT}\rvert/\lvert Pax_{all}\rvert$ |

## 4.4 City Characteristic Indicators

This study establishes a comprehensive framework to analyze city characteristics through three key dimensions:

**Population and City Scale** Population ($P$) serves as the primary classifier for city scale:

$$\text{City Scale} = \begin{cases} \text{Large}, & P > 200,000 \\ \text{Medium}, & 100,000 \leq P \leq 200,000 \\ \text{Small}, & P < 100,000 \end{cases}$$

This classification provides the fundamental basis for comparing SAV deployment strategies across different urban contexts.

**Network Structure Indicators** These metrics characterize the urban road network topology, the road networks have been transformed into simple graph $G(V, E)$:

1. **Link Density** ($\rho_L$): Measures network coverage intensity

$$\rho_L = \frac{|E|}{A} \tag{4.1}$$

where $|E|$ is the number of road segments and $A$ is the urban area. Higher density indicates better network coverage and more routing options.

2. **Network Connectivity** ($C$): also known as the Gamma Index ($\gamma$). This metric quantifies the ratio of existing edges to the maximum possible edges in a planar graph, assessing the network's overall interconnectedness. It is calculated as:

$$C = \frac{|E|}{3|V| - 6} \tag{4.2}$$

A higher value of $C$, approaching 1, indicates a more densely connected network with greater accessibility and more routing options, which can be advantageous for efficient SAV deployment and operations. Conversely, values closer to 0 suggest a sparser network structure(Rodrigue, 2024).

3. **Average Clustering Coefficient ($\bar{C}$)**: A measure of network transitivity, this indicates the overall tendency for nodes to form local clusters. It is the average of local clustering coefficients ($C_i$), where $C_i$ quantifies how close a node's neighbors are to forming a clique.

$$C_i = \frac{2T_i}{d_i(d_i - 1)} \tag{4.3}$$

where $T_i$ is the number of triangles passing through node $i$, and $d_i$ is the degree of node $i$. If $d_i < 2$, $C_i$ is defined as 0.

The average clustering coefficient is then:

$$\bar{C} = \frac{1}{|V|} \sum_{i \in V} C_i \tag{4.4}$$

A higher average clustering coefficient (closer to 1) indicates that, on average, the neighbors of nodes are densely interconnected, signifying a strong presence of local clustered structures within the network(Rodrigue, 2024).

4. **Network Meshedness** ($M$): This index, also known as the Alpha Index ($\alpha$) for planar networks, quantifies the degree of circuit redundancy or cyclicity within the network. It is calculated as the ratio of the actual number of fundamental circuits (or cycles) to the maximum possible number of circuits in a planar graph with the same number of nodes(Rodrigue, 2024). It is calculated as:

$$M = \frac{|E| - |V| + 1}{2|V| - 5} \tag{4.5}$$

A higher $M$ indicates a more grid-like topology with greater routing flexibility and redundancy, potentially benefiting SAV dispatching and resilience.

5. **Average Node Degree** ($\bar{d}$): This metric represents the average number of edges incident to a node (intersection) in the network. It provides a measure of the local connectivity at intersections.It is calculated as:

$$\bar{d} = \frac{1}{|V|} \sum_{i \in V} d_i \tag{4.6}$$

A higher $\bar{d}$ implies more immediate directional choices at intersections, offering increased local routing options. However, overall network navigability also depends on global topological properties.

**Demand Pattern Indicators** These metrics reveal travel demand characteristics:

1. **Commuting Trip Ratio** ($r_c$): Measures proportion of commuting trips

$$r_c = \frac{|Q_c|}{|Q|} \tag{4.7}$$

where $Q_c$ represents commuting trips and $Q$ total trips. Higher ratio indicates stronger commuting patterns.

2. **Average Commuting Distance** ($\bar{d_c}$): Reflects urban sprawl

$$\bar{d_c} = \frac{1}{|Q_c|} \sum_{i \in Q_c} d_i \tag{4.8}$$

Longer distances suggest greater potential for ride-sharing benefits.

These metrics collectively characterize the urban environment in which SAVs operate, enabling systematic analysis of fleet size optimization across different city scales.

## 4.5 Simulation Design

The simulation design of this study is structured into two levels: parameter calibration and sensitivity analysis simulation. This multi-level analytical approach ensures the setting of reasonable default parameters, making the system's performance align as closely as possible with reality or expected outcomes, while also evaluating the system's performance under different operational strategies.

### 4.5.1 Parameter Calibration Methods

Before conducting simulation, it is necessary to establish reasonable default values for the key parameters in the model. Parameters $\omega$, $p_{init}$ (initial fleet proportion to the number of demand), $c_w$ (vehicle waiting penalty factor), and $\alpha$ (weight factor for operators' utility) lack of direct real-world reference for their values, which makes it an essential task to determine the baseline combination through parameter analysis. To accomplish this, a specific calibration criterion is established for each parameter, stated as follows:

$\omega$ is a critical parameter for determining whether passengers should be considered for ride-sharing. The specific usecase is derived from ExMAS, and thus the utility formula for assessing whether a passenger is to join a shared ride is as follows:

$$\Delta U_i = \left(\pi \cdot \frac{l_i}{1000} \cdot \delta\right) + \beta_{ivt}\left(t_i - \omega_i(\hat{t}_{i,r} + |\Delta t^p_{i,r}|)\right) > 0 \tag{4.9}$$

The utility of shared travel for passenger $i$, denoted as $\Delta U_i$, represents the net benefit derived from choosing shared travel over non-shared travel, with a positive value indicating that shared travel is attractive. The base fare rate, $\pi$, is expressed in euros per kilometer, while the travel distance for passenger $i$, $l_i$, is measured in meters. A fare discount rate for shared travel, $\delta$, reduces the cost compared to non-shared travel. The value of time, $\beta_{ivt}$, reflects the cost passenger $i$ associates with both in-vehicle time and any delays. The direct travel time for non-shared travel is denoted by $t_i$, whereas $\hat{t}_{i,r}$ represents the in-vehicle travel time for passenger $i$ in shared ride $r$. Additionally, $|\Delta t^p_{i,r}|$ captures the absolute value of the pickup delay caused by shared ride $r$. The WtSR factor, $\omega_i$, indicates passenger $i$'s sensitivity to additional time costs due to sharing, with a smaller value signifying greater willingness to share.

The core of this formula lies in comparing the utility gain from fare discounts ($\pi(l_i/1000)\delta$) with the utility loss due to the increased total time cost caused by sharing (adjusted by the WtSR factor $\omega_i$ for in-vehicle time and delay time costs). Only when the utility gain exceeds the utility loss (i.e., $\Delta U_i > 0$) is the passenger considered willing to participate in the shared ride.

Given its direct impact on the generation of ride-sharing combinations, to ensure it reflects real-world conditions, $\omega$ is calibrated individually for each city using the actual ride-sharing rates from the ODiN dataset. The specific method involves setting a $\omega$ testing range with a step size of 0.02 and identifying $\omega$ value that results in a ride-sharing rate closest to the real rate among all test results. The metric used here is the Shared Ratio, which is the ride-sharing rate before vehicle assignment intervention, after the ride-sharing scheme is generated. This approach is chosen because, during the ride generation phase, the utility calculation involving $\omega$ directly reflects passengers' willingness to share rides without being influenced by other parameters. However, since this study considers the benefits to fleet operators and assumptions about fleet assignment strategies during vehicle assignment, the actual ride-sharing rate, Pooling Ratio, calculated post-assignment is affected by multiple other parameters. Therefore, using the Pooling Ratio to calibrate $\omega$ parameter lacks generality and robustness.

For the remaining operational parameters—$p_{init}$, $c_w$, and $\alpha$—a sequential calibration process is used to establish a robust baseline. The calibration process is guided by a set of assumptions to ensure the resulting baseline is transparent and consistent across all multi-city comparisons. The specific assumptions and calibration method of each key parameter are as follows:

- A city-specific **Initial Proportion ($p_{init}$):** calibrated to the theoretical minimum required to serve all demand. This methodological choice ensures a transparent baseline for multi-city comparisons and maximizes the system's sensitivity to other parameter changes. The criterion is to select the smallest value for which the number of new vehicles triggered during the simulation falls to zero.
- A globally applied **Waiting Cost ($c_w$,):** calibrated based on the assumption that the total cost of vehicle waiting should be comparable in magnitude to the total cost of driving. The criterion is to identify the parameter value at which the total vehicle waiting cost becomes comparable in scale to the total vehicle driving cost.

- A globally applied **Utility Balance ($\alpha$):** calibrated based on the strategic assumption that user and operator costs must be mutually constrained to maintain the system's attractiveness to users, the calibration aims to find a point of diminishing returns. The criterion is to select a weighting factor from the sensitivity analysis where further emphasis on operator cost yields minimal efficiency gains, while passenger costs begin to rise disproportionately.

### 4.5.2 City Characteristics Analysis Framework

The experimental plan to address the first research question (1) is designed as a comprehensive, multi-step comparative analysis across the 37 selected Dutch municipalities. This design focuses on isolating the impact of inherent urban characteristics on SAV system performance by maintaining a controlled operational environment. The experiment proceeds as follows:

First, a baseline simulation is executed for each of the 37 cities. Each simulation utilizes the city-specific demand profile and the full set of default parameters established through the calibration process described previously. This ensures a consistent and controlled operational baseline across all municipalities, allowing for a direct comparison of outcomes. The output of this process is a comprehensive set of KPIs for each city, which serves as the data foundation for the subsequent analyses.

Subsequently, a qualitative analysis is conducted to identify macro-level trends. The 37 cities are stratified into three categories—small, medium, and large—based on the population thresholds defined in the analysis framework (Section 4.4). Key performance indicators are then visualized using boxplots to compare their distributions across these three groups. This step aims to reveal initial patterns and significant performance differences associated with city scale.

Finally, a quantitative correlation analysis is conducted to uncover the statistical relationships between specific urban characteristic indicators and the resulting KPIs. This analysis is performed on two levels:

- **Global Analysis:** A correlation analysis is first performed on the entire dataset of 37 cities. This helps to identify strong, overarching relationships, particularly those driven by population scale, and to assess potential multicollinearity among the urban indicators themselves.
- **Group-Specific Analysis:** The correlation analysis is then repeated independently for each of the three city-size groups. By examining cities of similar population scale, this approach effectively controls for the dominant influence of population. It allows for the identification of more nuanced relationships, revealing which specific network or demand characteristics are most influential on SAV system performance within small, medium, or large urban contexts.

### 4.5.3 Analysis of Operational Strategies

This section primarily focuses on the system's sensitivity to changes in key parameters and their impact on system behavior. The approach first involves evaluating a broad range of parameter intervals to identify sensitive areas and understand the system's stability and adaptability under various operational strategies. Subsequently, a more focused analysis is conducted within these sensitive ranges to provide detailed insights. The analysis is conducted across cities of varying scales, aiming to explore how parameter sensitivity varies with city size and to provide empirical support for fleet operation strategies tailored to different urban scales.

The sensitivity analysis specifically examines the following parameters:

- **Maximum Pickup Delay($\Delta t^{p,\max}$)**: To investigate a key dimension of the main research question—the balance between operator cost and user service level—this study analyzes the impact of $\Delta t^{p,\max}$. This parameter functions as a hard constraint in the vehicle assignment algorithm, filtering out any potential solutions that violate the passenger waiting time limit. It therefore directly controls the trade-off between user service quality and operational efficiency. In the simulation, this parameter was varied from 100 to 1900 seconds, with a step size of 200 seconds.
- **WtSR and Shared Discount($\omega$ and $\delta$)**: Given that passengers' willingness to share ($\omega$) and the fare discount offered for shared rides ($\delta$) collectively determine the attractiveness of shared mobility and directly influence the core system metric of ride-sharing rate, this study conducts a joint sensitivity analysis of these two parameters. By employing a grid search approach, their interaction effect on system performance, particularly on ride-sharing behavior and associated efficiency metrics, is systematically evaluated. This analysis is designed to directly address research

sub-question 2. The grid search explored $\omega$ values from 0.9 to 1.1 with a step of 0.02, and $\delta$ values from 0.1 to 0.9 with a step of 0.1.

- **Scale Factor($s$)**: First of all, by systematically varying $s$, this analysis explores whether conclusions drawn from a 1% sampled demand remain applicable under different demand scales. This helps confirm the model's consistency in real-world operations. Second, by incrementally increasing demand in cities of different sizes (medium and small), this study investigates whether key performance indicators exhibit significant patterns or potential economies of scale. This is crucial for understanding the potential and challenges of scaling SAV services across diverse urban contexts. Additionally, evaluating system performance under different scale factors provides operators with guidance on resource allocation to optimize operational efficiency while meeting demand across varying scales. This analysis directly addresses research sub-question 3. The analysis of the scale factor was performed in two parts: a direct sensitivity analysis varied $s$ from 0.01 to 0.10 with a step size of 0.01. Subsequently, a joint analysis with passenger resistance was conducted, where $s$ varied over the same range while $\omega$ varied from 1.02 to 1.12 with a step of 0.02.

During the sensitivity analysis, parameters analyzed individually are varied within their respective test ranges, while other parameters are held at their default or calibrated values. For parameter pairs under joint analysis, exploration is conducted simultaneously within a two-dimensional parameter space.

### 4.5.4 Public Transport Competition analysis

When large-scale deployment of SAVs is implemented as a solution for urban mobility, the relationship between SAVs and conventional public transport should be reconsidered. This section focuses on the competitive relationship between SAVs and traditional rail public transport, examining the interactions between these two modes under varying conditions. Two core aspects are addressed in this study:

- Public transport pricing: If public transport discount is applied, how can SAVs maintain their service attractiveness?
- Impact of traffic congestion: How does the competitive relationship between SAVs and public transport change under different levels of traffic congestion?

To address these questions, the simulation in this section will map the two aspects to key model parameters: $v_{avg}$ is used to simulate the degree of traffic congestion, and $\delta_{PT}$ is applied to adjust public transport fares. Two groups of simulation are designed to systematically quantify the effects of each factor.

The first scenario explores the impact of different $\delta_{PT}$ on the competitive relationship between SAVs and public transport, as well as the indirect effects on the SAV system. Here, the $v_{avg}$ is set to a baseline of 8 m/s, while the $\delta_{PT}$ parameter is varied from 0.1 to 0.9 in increments of 0.1.

A supplementary scenario further examines the joint effects of traffic congestion and public transport pricing using a grid search approach. In this setting, the $\delta_{PT}$ is varied from 0% to 90% in 10% increments, and the $v_{avg}$ ranges from 5 to 10 m/s in unit increments. This design enables a comprehensive assessment of the combined influence of congestion and fare adjustments on the competition structure between SAVs and public transport.

To model the impact of congestion, the analysis focuses on the morning peak hour from 8:00 to 9:00. Specifically, the simulation for this analysis is confined to a one-hour duration, using only the travel demand generated between 8:00 and 9:00, with demand from other periods excluded. This period is selected as it exhibits the maximum travel demand density in the dataset. It is therefore assumed that this timeframe corresponds to the most significant levels of road network congestion, making it the most critical for analyzing shifts in mode choice under traffic pressure. Within this context, the average operating speed of SAVs ($v_{avg}$) is treated as a key variable parameter. By systematically reducing this speed, different levels of road network congestion and their negative impacts on SAV operational efficiency can be simulated. Meanwhile, the speed of public transport is held constant, based on the assumption that the rail-based services considered in this study operate on dedicated infrastructure and are therefore unaffected by road traffic congestion.

All public transport related parameters, including $\lambda_{SAV}$, $v_{PT}$ and price-related paramters are defined in detail in Appendix C. A key parameter in this model is the SAV nest scale parameter, $\lambda_{SAV}$, which governs the perceived substitutability between shared and non-shared SAV options. For the primary experiments detailed in this study, a baseline value is adopted for $\lambda_{SAV}$ to maintain focus on the core

research questions. To ensure that the main conclusions are robust and not artifacts of this specific parameter choice, a dedicated sensitivity analysis was conducted on $\lambda_{SAV}$. This analysis confirms the stability of the model's macroscopic predictions under different $\lambda_{SAV}$ values. The detailed methodology and results of this robustness check are presented in Appendix K.

# Chapter 5

# Simulation Results and Analysis

This chapter presents the simulation results and analysis. It first establishes a baseline for key model parameters through calibration. The subsequent analysis examines the impacts of urban heterogeneity, $\Delta t^{p,\max}$, the interaction between $\omega$ and $\delta$, and demand scale. The chapter concludes by first defining the competitive structure between SAV and public transport across different travel distances, and then examining how this structure is reshaped by pricing strategies and traffic congestion.

## 5.1 Calibration of Default Values for Key System Parameters

The calibration of each parameters follows method illustrated in Section4.5.1. The calibrated baseline combination directly enables the investigation of research sub-question 1, which examines the relationship between urban characteristics and SAV system performance, by ensuring that all inter-city comparisons are made under a consistent and well-justified set of operational assumptions.

### 5.1.1 Calibration of Willingness-to-Share Resistance Factor($\omega$)

To ensure that the simulation results closely reflect reality and to provide a reliable baseline for subsequent sensitivity analysis and policy evaluation, this study first calibrates the key parameter $\omega$. $\omega$ parameter reflects passengers' tolerance for the additional time costs associated with shared rides, directly influencing the generation of potential shared trips and the overall ride-sharing potential of the system.

During the calibration process, for each studied city, a fixed range of $\omega$ test values (ranging from 0.8 to 1.3, with a step size of 0.01) was used. These simulated proportions are then compared with the actual proportion of shared trips calculated from the ODiN dataset for that city (Real Ratio in the figure). $\omega$ value that results in the closest match between the simulated and actual proportions is selected as the calibrated $\omega$ value for that city.

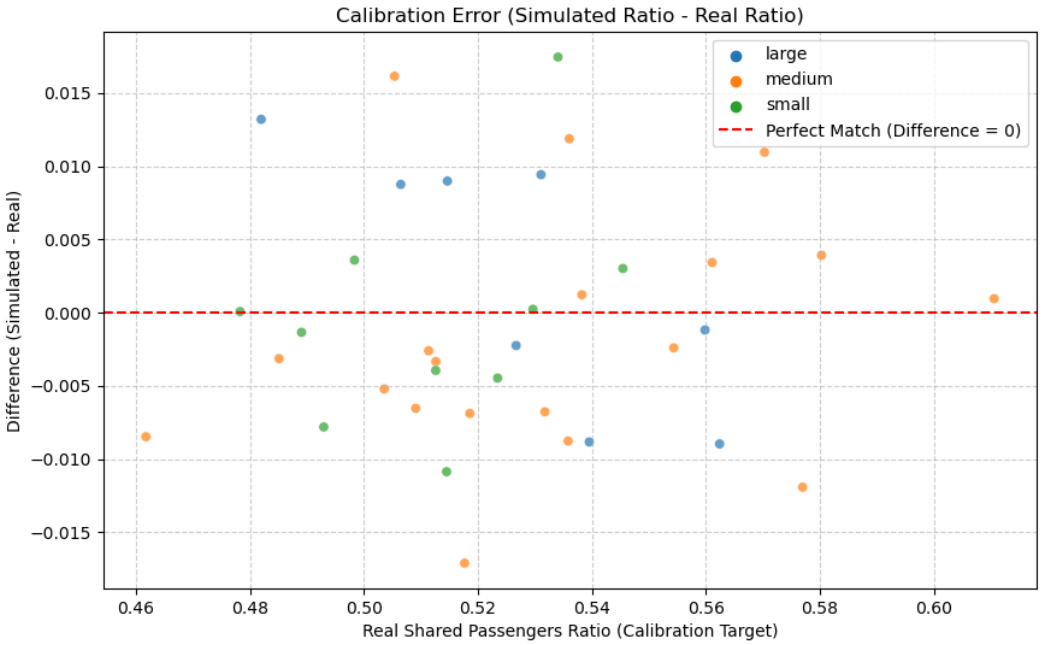



Figure 5.1: Deviation Between Calibrated and Actual Proportions of Shared Trips

Through this approach, this study is able to determine individual $\omega$ baseline values for all 37 studied cities. The calibration results, as shown in Figure 5.1, indicate that this method reduces the difference between the simulated and actual shared trip rates to within 2%. The calibrated $\omega$ values vary across cities (ranging approximately from 0.8 to 1.06). Within the framework of this study, these variations do not necessarily reflect differences in residents' subjective willingness to share rides. Rather, they primarily represent the necessary adjustments to align the model's endogenous ride-sharing potential with observed real-world values across cities with distinct road network structures and demand characteristics. Detailed calibration results, including $\omega$ values for each city, corresponding population figures, and the actual versus simulated shared trip proportions used in the calibration process, are provided in Appendix.

Table 5.1: Statistics of Actual Shared Ratio and Best $\omega$ by City Scale

| **City Scale** | **Real Shared Ratio** | | **Best $\omega$** | | **Count** |
|---|---|---|---|---|---|
| | **Mean** | **Std Dev** | **Mean** | **Std Dev** | |
| large | 0.5278 | 0.0270 | 1.0300 | 0.0256 | 8 |
| medium | 0.5327 | 0.0363 | 0.9663 | 0.0315 | 19 |
| small | 0.5118 | 0.0218 | 0.9220 | 0.0601 | 10 |

Table 5.1 provides the statistical analysis of the shared trip ratio and $\omega$ values across different city scales. It is apparent that $\omega$ values for small cities exhibit a larger standard deviation, whereas the standard deviation of the shared trip ratio in small cities is lower than that in medium-sized cities. This indicates that the calibrated $\omega$ values effectively capture the critical parametric differences that drive varying levels of ride-sharing across cities. Small cities may have significant heterogeneity in dimensions that are already captured by the model, such as demand patterns and network topology. This inherent diversity is the primary reason for the significant variation in calibrated $\omega$ values among small cities, even though their final ride-sharing rates appear to be similar.

### 5.1.2 Calibration of Initial Proportion ($p_{init}$)

The initial fleet size for each simulation is determined by $p_{init}$ parameter, which calculates the number of vehicles as a proportion of the total travel demand. The primary objective in setting this parameter is to establish a fleet size that is both resource-efficient and sufficient to serve all requests without needing

to trigger the "new vehicle" mechanism. Activating this mechanism introduces vehicles with fixed, predetermined costs and pickup times (as per Equation 3.8), which could bias the simulation results. To determine the appropriate $p_{init}$ for each municipality, a sensitivity analysis was conducted. The key metric for this calibration was the number of new vehicles triggered during the simulation. The optimal value was defined as the smallest proportion at which the number of triggered vehicles dropped to zero.

Table 5.2: Optimal $p_{init}$ Statistics by City Size

| City Size | Mean | Std. Dev. | Count |
|---|---|---|---|
| Large | 8.38 | 0.51 | 8 |
| Medium | 9.11 | 2.02 | 19 |
| Small | 10.4 | 1.71 | 10 |

These individually calibrated values are used as the default settings in all subsequent experiments to ensure a consistent and unbiased baseline. A summary of the calibrated values by city size is presented in Table 5.2. The result revealed a consistent trend across city scales: smaller cities require a proportionally larger initial fleet relative to their demand volume to meet this criterion. This suggests that operators in smaller markets may face different relative fleet-sizing challenges compared to those in larger cities.

A detailed description of the calibration methodology, including the analysis of representative cities, all supporting figures, and an extended analysis of system performance under varying fleet abundance, is provided in Appendix J.1.

### 5.1.3 Calibration of Waiting Cost ($c_w$)

The parameter $c_w$ (€/second) penalizes inefficient vehicle idle time. To determine an appropriate value, a sensitivity analysis was conducted. Detailed analysis on a representative city (see Appendix J.2 for figures) reveals a critical trade-off: if the penalty is too low, vehicles wait excessively, which is inefficient; if the penalty is too high, the system avoids waiting by forcing vehicles into long, costly pickup trips, which paradoxically increases overall operational costs and can worsen passenger service.

As outlined in the calibration criteria, the goal is to find a balance point for $c_w$ penalty. This balance is crucial: the penalty for inefficient idle time must be significant enough to influence scheduling, but not so high that it forces vehicles into excessively long and costly pickup trips just to avoid minor waits. Therefore, the specific criterion is to identify the parameter value where the total vehicle waiting cost becomes comparable in scale to the total driving cost.

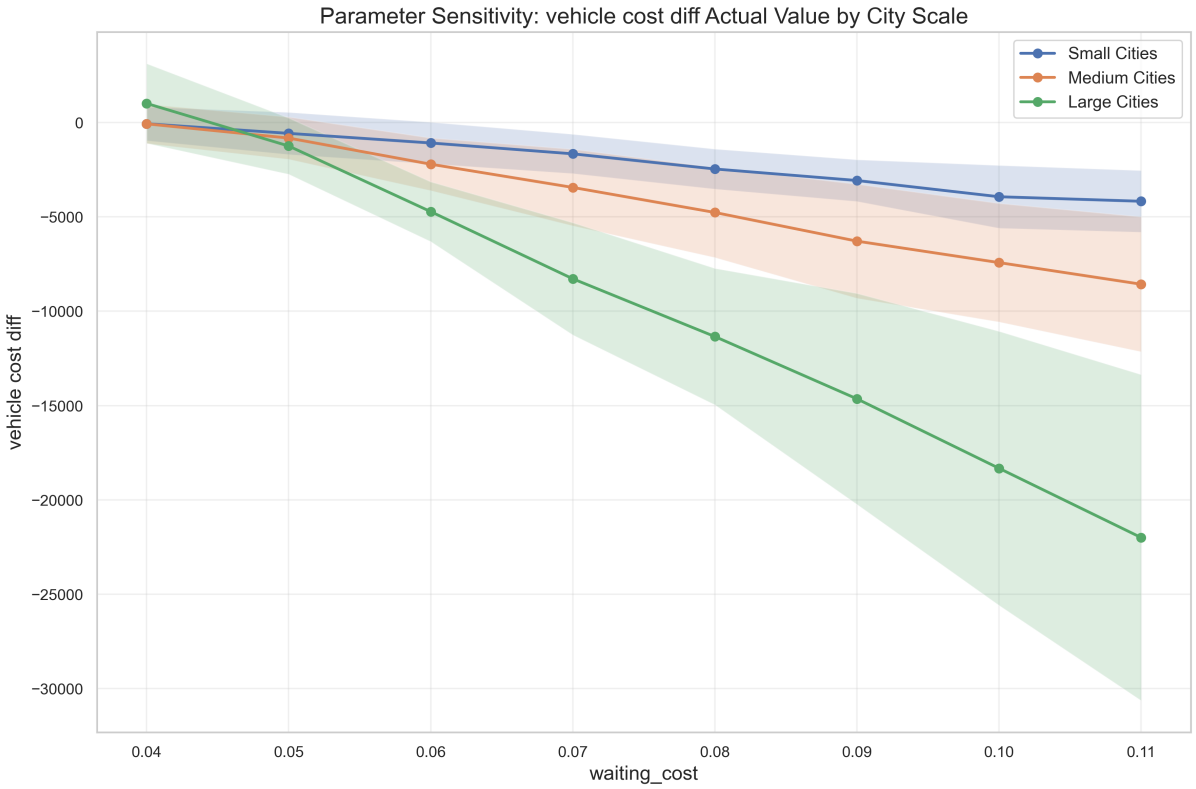


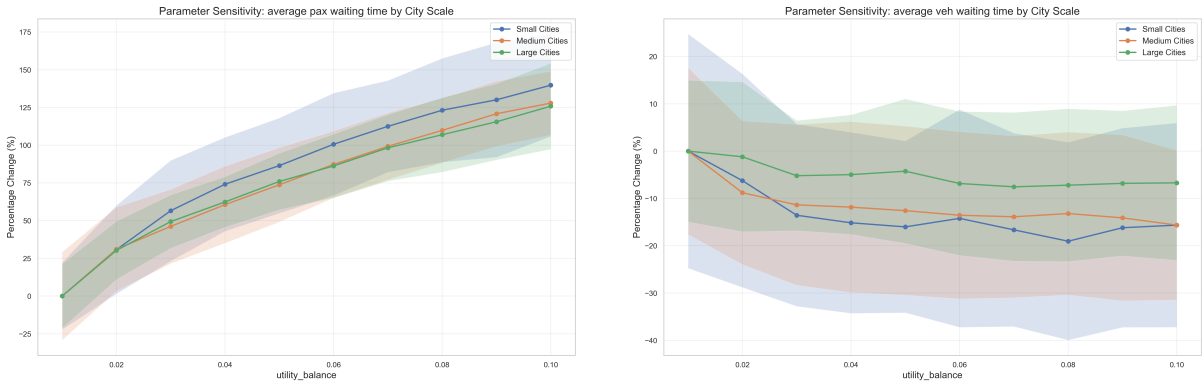

Figure 5.2: Difference Between Vehicle Driving Cost and Vehicle Waiting Cost

Figure 5.2 shows the difference between vehicle driving cost and vehicle waiting cost. A negative value indicates that the vehicle waiting cost has exceeded the vehicle driving cost. The balance point between vehicle driving and waiting costs for most cities is around 0.05. Within the tested range, *waiting cost = 0.05* provides a well-balanced point. Furthermore, this value also makes the vehicle waiting cost and travel cost comparable in scale, which helps to balance the objective function. Therefore, selecting 0.05 as the default value aims to effectively suppress vehicle idle waiting while avoiding potential negative chain effects caused by setting the parameter too high, ensuring that passenger service levels and overall system operational efficiency are maintained at a good state.

### 5.1.4 Calibration of Utility Balance ($\alpha$)

The operation of a sustainable SAV system requires balancing operator efficiency with user attractiveness. This study models this challenge through a $\alpha$ parameter, which weights the relative importance of operator costs versus passenger utility in the optimization function. While the detailed calibration process for this parameter is provided in Appendix J.3, the analysis itself revealed critical strategic insights that constitute a key finding of this research.

The primary finding is the direct trade-off between operational efficiency and service quality. As shown in Figure 5.3, increasing the focus on operator costs (a higher $\alpha$) effectively reduces vehicle waiting time, but at the direct expense of increased passenger waiting time. The analysis identified a turning point around a value of **0.06**. At this point, the rate of decrease in vehicle costs begins to slow, while the increase in passenger costs stabilizes. This value represents a data-supported balance point, achieving significant operational efficiencies without excessively penalizing the user experience. For this reason, it is adopted as the default value in subsequent analyses.




(a) Passenger Waiting Time (b) Vehicle Waiting Time

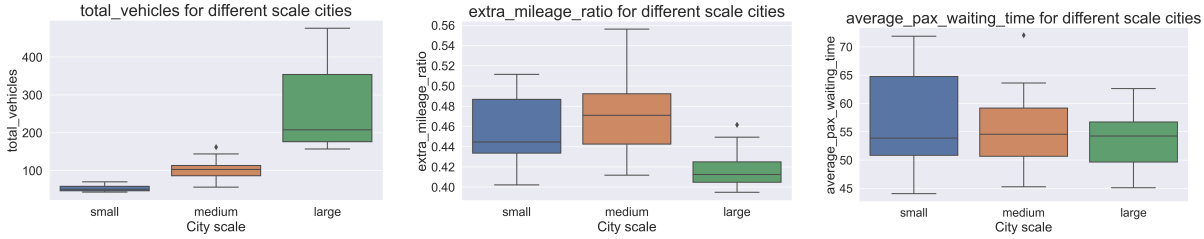
Figure 5.3: Impact of $\alpha$ on Passenger and Vehicle Waiting Times

## 5.2 Correlation Analysis between City Characteristics and Operational Indicators

The determination of baseline parameters establishes the foundation for multi-city simulation experiments. This section directly addresses the first research sub-question 1, exploring the correlation between city characteristics and operational performance indicators across different city scales.

### 5.2.1 Qualitative Group Comparison

Under the default parameter settings, a comparative analysis of KPIs was performed across different city size groups (large, medium, and small cities). This analysis aims to initially reveal the overall trends and general patterns of how city scale affects SAV system performance.




(a) Total Vehicles (b) Extra Mileage Ratio (c) Average Passenger Waiting Time

Figure 5.4: Comparison of KPIs by City Scale using Boxplots.

After simulating different scale cities using the preset baseline parameter combination, a comparative analysis of the performance of various KPIs across large, medium, and small city groups was conducted. This analysis aims to preliminarily reveal the macroscopic impact patterns of city scale effects on the overall operational characteristics and various performance indicators of the SAV system.

Regarding fleet size indicators, a clear scaling effect was observed. The median values for both the required total number of vehicles (*total vehicles*) show a significant stepwise increase with city size. Large cities require more resources than small and medium-sized cities, and the data variance also increases with scale. These results align with basic expectations that larger demand volumes need greater vehicle resource inputs, which usually implies higher overall operational investment.

From the perspective of vehicle operational efficiency, large cities show better performance. Specifically, the (*extra mileage ratio*) in large cities are significantly lower than that in medium and small cities. This phenomenon might mainly caused by the higher population density in large cities, which allows vehicles to find new matches in nearby areas easily after completing a service, effectively reducing deadhead travel distance. Still, specific data to support this assumption are required to prove its correctness, which will be shown in the following quantitative analysis.

Regarding the service level, *average pax waiting time* shows comparable median values across the different city scales, but a notable distinction emerges in the consistency of service. As illustrated by the boxplot, small cities display a considerably wider distribution of waiting times. This suggests that the service level in smaller cities is less predictable, with passengers potentially facing a greater variability in their waiting experience compared to the more stable and consistent service observed in medium and large cities.

### 5.2.2 Quantitative Indicators Correlation Analysis

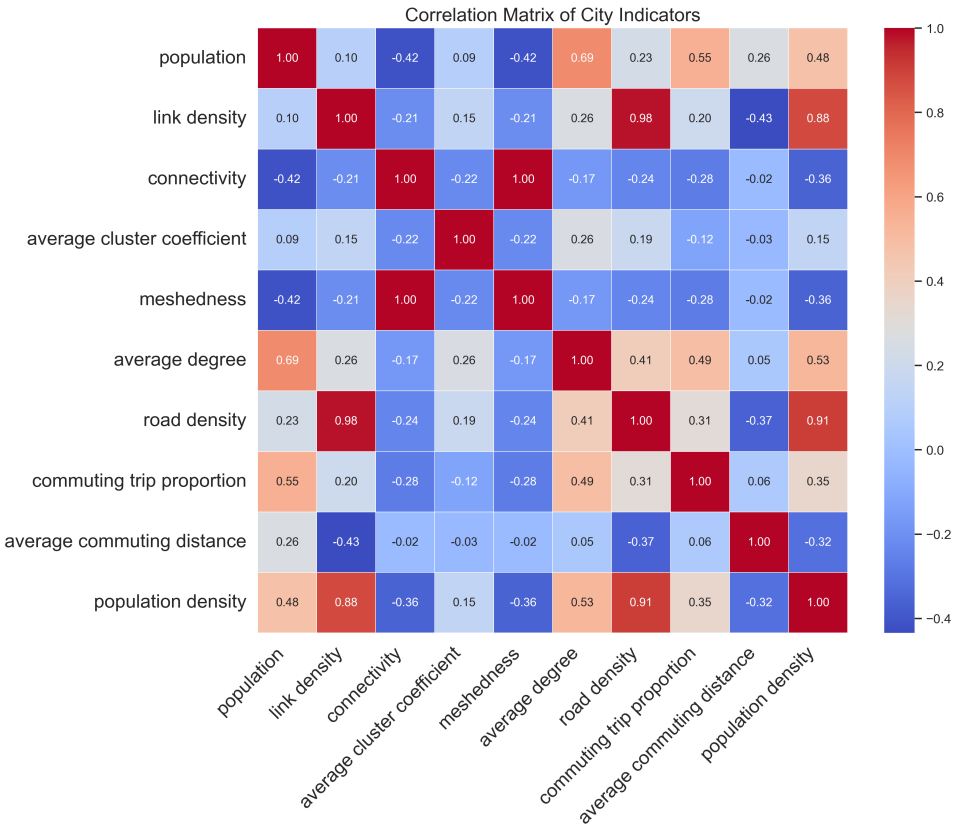
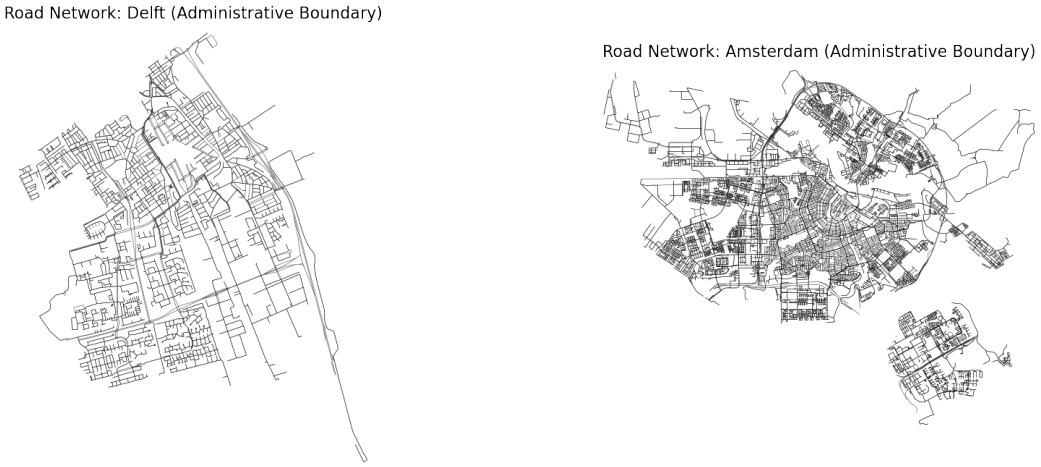



Figure 5.5: Correlation Matrix Heatmap between Urban/Demand Characteristic Indicators

Before delving into the relationship between city characteristic indicators and KPIs, it is essential to first examine the internal correlations among these indicators themselves(Figure 5.5). As previously mentioned, population size (*population*) exhibits a strong positive correlation with other indicators, such as *link density*, *road density*, *population density* and some network structure indicators like *average degree*. This implies that when these indicators are observed to correlate with a certain KPI, it might primarily be an indirect result of the population scale effect. Therefore, the analysis in this section will focus more on identifying city characteristics that can independently explain variations in KPIs after controlling for the influence of population size, as well as correlations that reveal non-obvious mechanisms.

Further examination of network density-related indicators reveals a strong positive correlation among *link density*, *road density*, and *population density*. For instance, the correlation coefficient between *link density* and *road density* is 0.98, while their respective correlations with *population density* are also substantial, at 0.88 and 0.91. The strong intercorrelation among these density indicators suggests they collectively reflect a broader concept of urban density or compactness. This understanding is crucial for the subsequent analysis of relationships between urban characteristic indicators and KPIs, as it helps in recognizing potential multicollinearity. Acknowledging these strong intercorrelations allows for a more nuanced interpretation of how each density-related characteristic might influence KPIs, by considering that the observed effect of one density indicator could be partly attributable to its close association with others. This insight is vital for discerning more direct influences from confounded ones when evaluating the impact of urban form on system performance.

Additionally, the correlation coefficient between *connectivity* and *meshedness* approaches 1.0 in the all-city scenario, indicating their high collinearity when applied to the urban road networks analyzed in this study. This strong linear relationship suggests that, in this specific analytical context, they are capturing highly similar underlying dimensions of network attributes. Consistent with the defination of Rodrigue (2024) on related network metrics, the *meshedness* used in this study's model corresponds to the Alpha Index ($\alpha$), represents the richness of circuits in the network. *Connectivity* corresponds to the Gamma Index ($\gamma$), represents the density of connections within the network. Their calculation formulas are detailed in Section 4.4. It is noteworthy that both indicators exhibit a moderate negative correlation with *population*. This implies that within the urban samples of this study, cities with larger population scales may possess lower overall network connectivity or meshedness. This phenomenon that where larger cities exhibit reduced overall road network connectivity and circuit redundancy, appears counterintuitive.




(a) The road network of Delft

(b) The road network of Amsterdam

Figure 5.6: Comparison of Road Network Structures: Delft and Amsterdam

It might reflect the definition standard of administrative boundaries for large cities, which might potentially cover more suburban areas that characterized by less developed or lower-density road network structures. In contrast, smaller cities might have their road network boundaries focused more on well-developed and homogeneous areas, potentially resulting in higher average values for these indicators. Figure 5.6a and 5.6b shows the road network within the administrative boundaries of Amsterdam and Delft, it explains why the performance of large cities on these indicators would not necessarily represents inefficiency in their central area road networks, but rather reflect the complexity and uneven development of large cities' spatial structures and the homogeneous network form of smaller cities.

Furthermore, regarding travel demand characteristics, the positive correlation (0.55) between *commuting trip proportion* and *population* might indicate that large cities often have a higher proportion of employed population. Consequently, when analyzing the specific impacts of commuting patterns on SAV systems, the underlying influence of population scale must be considered.

In summary, a systematic review of the intercorrelations among urban characteristic indicators is an essential step before proceeding with the analysis of their impacts on KPIs. This review not only reveals the complex interactions among multiple indicators related to *population*, laying the groundwork for controlling potential confounding variables and identifying true independent influencing factors in subsequent analyses, but also provides a clearer and more focused research path for in-depth investigation into how specific urban attributes shape the performance of SAV systems.

### 5.2.3 High Correlation Analysis between City Characteristic Indicators and KPIs

When developing and implementing SAV systems in different cities in real-world case, understanding how different city characteristics influence the system performance is crucial. This section conducts a comparative analysis of the correlations between city characteristic indicators and SAV system KPIs, both across all city samples(37 cities) and within distinctive three city size groups. The aim for analyzing all citiees as a group is to reveal the common patterns that may appear in all cities, along with performance differences that caused by scaling effects, and further propose plausible hypotheses for the observed correlations. Furthermore, to investigate how other urban features independently affect SAV system KPIs in cities of different population sacles, analyzing data separately for different city size groups can be considered an approximate method for controlling the population size variable, which is particularly relevant for medium and small cities, while separate analysis of large cities can more directly illustrate the impact of scaling effects that related to significant population differences.

Table 5.3: High Correlation Analysis between City Characteristic Indicators and KPIs across all cities

| **City Indicator** | **KPI** | **Correlation** | **P-Value** |
|---|---|---|---|
| population | peak concurrent vehicles | 0.9796 | 1.434e-24 |
| population | total vehicles | 0.9758 | 2.296e-23 |
| population density | average pickup time | -0.5111 | 1.704e-03 |
| average degree | peak concurrent vehicles | 0.7314 | 5.995e-07 |
| average degree | total vehicles | 0.6975 | 3.222e-06 |
| average degree | saved miles ratio | 0.5906 | 1.878e-04 |
| commuting trip proportion | total vehicles | 0.6005 | 1.370e-04 |
| commuting trip proportion | peak concurrent vehicles | 0.6003 | 1.378e-04 |
| average commuting distance | time utilization rate | 0.5388 | 8.391e-04 |
| road density | average pickup time | -0.5584 | 4.902e-04 |
| link density | average pickup time | -0.5268 | 1.150e-03 |

Table 5.4: Selected Correlations between City Indicators and KPIs in Small Cities (N=8)

| City Indicator | KPI | Correlation | P-Value |
|---|---|---|---|
| population density | avg rides per vehicle | 0.750 | 0.032 |
| average degree | saved miles ratio | 0.903 | 0.002 |
| average cluster coefficient | vehicle reuse rate | 0.839 | 0.009 |
| average commuting distance | total vehicles | 0.802 | 0.017 |

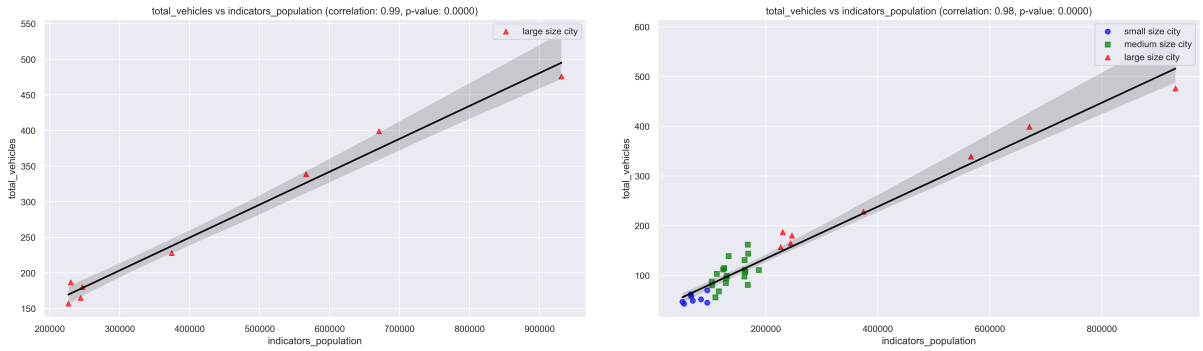
Table 5.5: Selected Correlations between City Indicators and KPIs in Medium Cities (N=19)

| City Indicator | KPI | Correlation | P-Value |
|---|---|---|---|
| population | peak concurrent vehicles | 0.703 | <0.001 |
| population | total vehicles | 0.542 | 0.017 |
| avg commuting distance | average pickup time | 0.509 | 0.026 |
| road density | average pickup time | -0.540 | 0.017 |
| link density | average pickup time | -0.523 | 0.022 |

Table 5.6: Selected Correlations between City Indicators and KPIs in Large Cities (N=8)

| City Indicator | KPI | Correlation | P-Value |
|---|---|---|---|
| population | total vehicles | 0.991 | <0.001 |
| population | peak concurrent vehicles | 0.989 | <0.001 |
| avg commuting distance | pooling ratio | 0.884 | 0.004 |
| avg commuting distance | peak concurrent vehicles | 0.839 | 0.009 |
| population density | peak concurrent vehicles | 0.849 | 0.008 |
| population density | total vehicles | 0.837 | 0.010 |
| average degree | pooling ratio | 0.807 | 0.016 |
| average degree | peak concurrent vehicles | 0.786 | 0.021 |
| average degree | saved miles ratio | 0.739 | 0.036 |

These analysis are based on the results presented in Table 5.3 (summary of correlations for all cities), Table 5.4 (correlations for small cities), Table 5.5 (correlations for medium-sized cities), and Table 5.6 (correlations for large cities), only significant correlations (P-Value $< 0.05$) with absolute value $> 0.5$ are shown. It should be noted that these tables do not include results related to any cost functions, as these are directly associated with the scale of total demand and total fleet size, thus offering limited comparative value. Conversely, key indicators that contribute to the cost structure, such as *average passenger waiting time* and *average vehicle waiting time*, along with fleet efficiency metrics like *vehicle reuse rate* and *average rides per vehicle*, are more meaningful for comparison as they are calculated based on average values.




(a) Total Vehicles vs. Population - Large Cities

(b) Total Vehicles vs. Population - All Cities

Figure 5.7: Relationship between Total Vehicle Requirements and City Population

To begin with, *population*, as the most direct indicator of city scale, plays a fundamental role in driving the demand volume of this research. In the total sample of 37 cities, *population* exhibits a strong positive correlation with both *peak concurrent vehicles* and *total vehicles*, intuitively reflecting the positive correlation between population scale and the required fleet size. Notably, in the large city group, the correlation between *population* and fleet size indicators (*total vehicles* correlation coefficient as 0.99, *peak concurrent vehicles* correlation coefficient also reaching 0.99) approaches a near-perfect level. As can be seen in Figure 5.7a and 5.7b, this phenomenon is closely related to the substantial variation in population scale within large cities group. In other words, the considerable range of population values within the large city group makes the population scaling effect more evident, almost linearly determining the fleet size. In contrast, in medium-sized cities, while the correlation between *population* and fleet size remains significant, the strength of this correlation weakens, suggesting that at this scale, other urban characteristics (such as urban form or diversity of travel patterns) may begin to have a more complex combined influence on fleet size. In the small city sample, however, no KPIs were found to have a significant high correlation with population scale, indicating that the similar population sizes in small cities allow for a relatively independent analysis of the impact of other city indicators on KPIs, effectively controlling for population scale to some extent.

Table 5.7: Descriptive Statistics of Average Commuting Distance and Population Density by City Size

| Indicator | City Size | Mean | Min | Max | Variance |
|---|---|---|---|---|---|
| *Average Commuting Distance(m)* | Small (N=8) | 4272.37 | 2984.32 | 5964.10 | 1.1519e+06 |
| | Medium (N=19) | 4060.99 | 2969.98 | 5275.00 | 5.2838e+05 |
| | Large (N=8) | 4504.53 | 3996.90 | 4961.08 | 9.5682e+04 |
| *Population Density($n/km^2$)* | Small (N=8) | 1964.32 | 180.40 | 4352.82 | 2.2404e+06 |
| | Medium (N=19) | 2063.72 | 455.31 | 5198.98 | 2.4232e+06 |
| | Large (N=8) | 3249.64 | 1517.59 | 5041.07 | 1.8033e+06 |

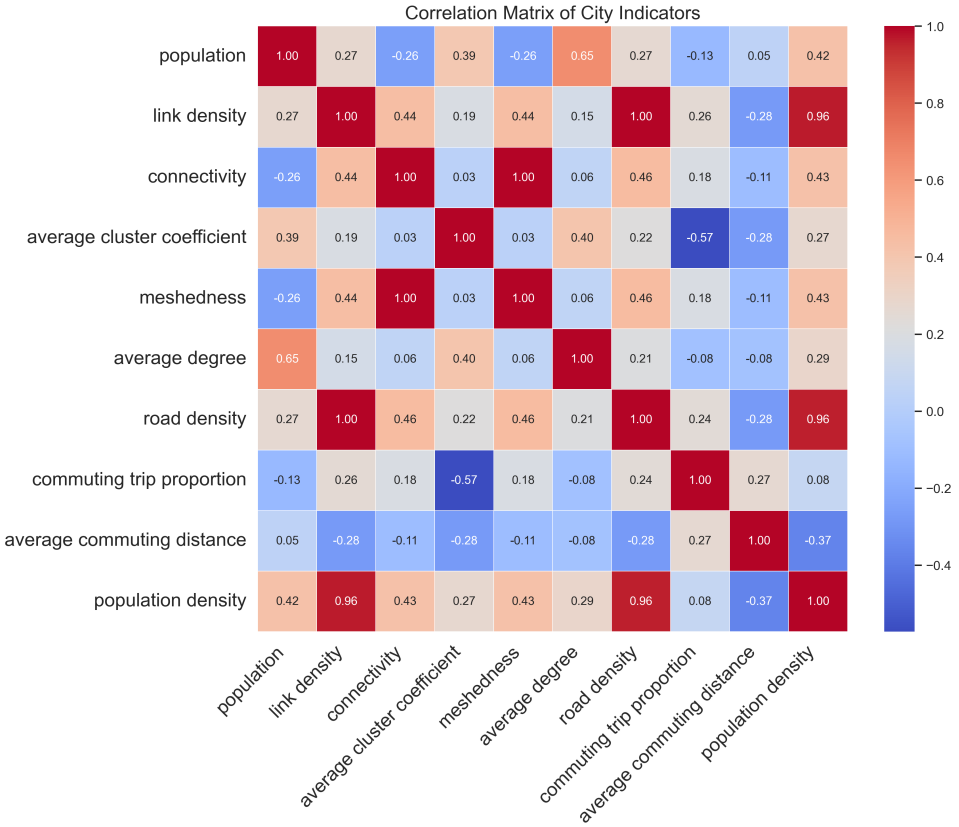



Figure 5.8: Small Cities City Indicators Correlation

Moreover, *average commuting distance*, as a key indicator reflecting travel demand characteristics, exhibits multidimensional impacts on this research's KPIs across different city scales. In the full city sample, *average commuting distance* shows a positive correlation with *time utilization rate* (it represents the proportion of time a vehicle is engaged in service, pickup, or waiting, out of its total active time). This suggests that longer commuting trips may provide vehicles with longer continuous service periods, thereby contributing to an overall improvement in vehicle time utilization efficiency. Furthermore, in small cities, *average commuting distance* presents a strong positive correlation (0.80) with *total vehicles*. As observed in Table 5.7, small cities have the highest variance in *average commuting distance*, indicating significant variability in this factor. Moreover, Figure 5.8 shows a fairly weak correlation (0.05) between *average commuting distance* and *population* in small cities. Therefore, the small city sample allows for a relatively independent analysis of the impact of *average commuting distance*. This could imply that longer average commuting distances lead to more dispersed travel patterns and longer vehicle occupation times, consequently requiring more vehicle resources to meet these long-distance travel demands, particularly in small cities where demand density might be relatively lower (see Table 5.7).

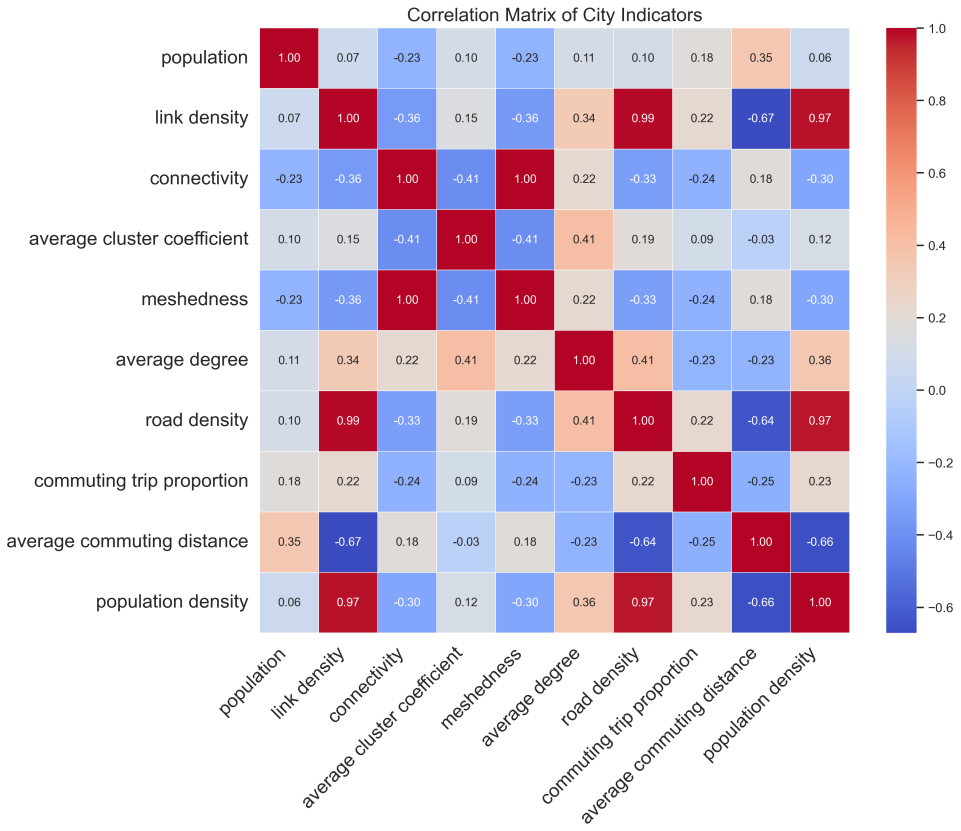



Figure 5.9: Medium Cities City Indicators Correlation

In medium-sized cities, *average commuting distance* shows a positive correlation(0.50) with *average pickup time*. However, as shown in Figure 5.9, the average commuting distance in medium-sized cities has a strong negative correlation (-0.84) with road density. Furthermore, road density and link density exhibit a stronger negative correlation with average pickup time(-0.54 and -0.52 respectively). Therefore, the impact on average pickup time here is more likely due to increased detours resulting from low road density, rather than an independent effect of average commuting distance.

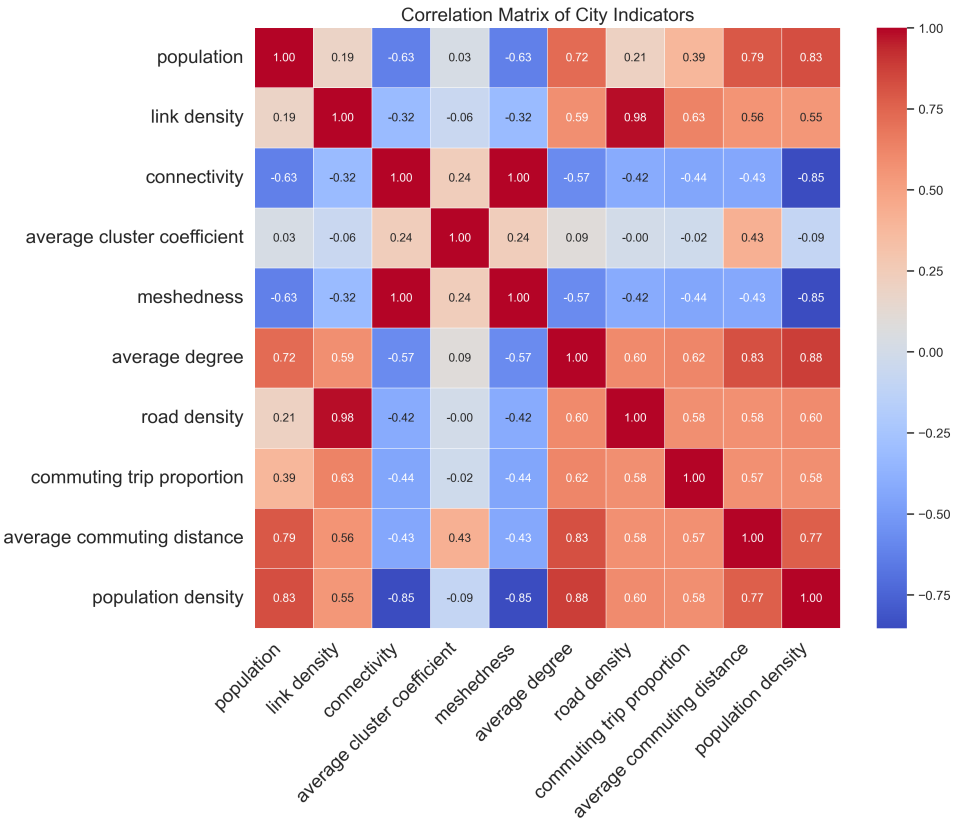



Figure 5.10: Large Cities City Indicators Correlation

In large cities, *average commuting distance* shows a strong positive correlation with both *pooling ratio* and *peak concurrent vehicles*. However, it is important to note that in the context of large cities, average commuting distance exhibits very strong positive correlations with many indicators due to the dominant role of population. Given that the *pooling ratio* does not have a significant correlation with population, it can be inferred that in large cities, the independent impact of average commuting distance is primarily reflected in its driving effect on the *pooling ratio*. From the positive correlation between *average degree* and *pooling ratio* (0.81), and that between two city characteristic indicators *average degree* and *average commuting distance* (0.83), it can be deduced that for the large city group with significant population differences, these two city characteristic indicators are the main contributors to the impact on the *pooling ratio*. The underlying reason may be that longer average commuting distances provide economic and time motivation for pooling, while a higher *average degree* offers convenient road network infrastructure for pooling, both of which jointly promote an increase in the *pooling ratio*. Conversely, *peak concurrent vehicles* shows a strong positive correlation with population, so the impact on *peak concurrent vehicles* here can be considered as a result of the combined effect of multiple factors dominated by population.

In summary, the impact pattern of *average commuting distance* varies with city scale, also reflecting the complexity of its joint effects with other indicators. For instance, in small cities, it may be in an isolated manner, directly related to the supply pressure on fleet resources, while in medium-sized cities, it combines with road density to affect vehicle service accessibility, in large cities, it combines with *average degree* to become a key factor for motivating pooling behavior.

Turning to the road network structure, as a core element affecting the operational efficiency of SAV systems, also shows that the impact of its specific indicators varies with city scale. The most relevant indicator is *average degree*, in the full city sample, this indicator shows a positive correlation with *peak concurrent vehicles*, *total vehicles*, and *saved miles ratio* (in this study, *saved miles ratio* represents the vehicle mileage saved due to efficient pooling route planning compared to the sum of the shortest paths for all passengers). Although *average degree* is positively correlated with *population* (0.69), no high correlation between population and saved mileage was found in any of the statistical data. In the data for both small and large cities, it can be observed that *saved miles ratio* has a high correlation with *average degree* but not with other city characteristic indicators (in medium-sized cities, the correlation coefficient between *average degree* and *saved miles ratio* is 0.49 with a p-value of 0.03, which can also be

considered a moderately significant correlation). Therefore, it can be inferred that the impact of *average degree* on *saved miles ratio* is relatively isolated. We can thereby hypothesize that a higher *average degree* implies more numerous connections between road network nodes, providing more path options for SAV dispatch, thus supporting the possibility of more efficient pooling combinations and facilitating mileage savings. As for its impact on the other two indicators, *peak concurrent vehicles* and *total vehicles*, it is more likely related to the direct increase in vehicle numbers brought about by population scale.

Table 5.8: Description Statistics for Average Cluster Coefficient by City Size

| City Size | Count | Mean | Std | Min | Max | Variance |
|---|---|---|---|---|---|---|
| Small | 8.0 | 0.069180 | 0.010557 | 0.056768 | 0.085188 | 0.000111 |
| Medium | 19.0 | 0.065277 | 0.007463 | 0.053794 | 0.083564 | 0.000056 |
| Large | 8.0 | 0.069068 | 0.008658 | 0.056173 | 0.082138 | 0.000075 |

Another road network structure indicator worthy of attention is *average cluster coefficient*. A unique finding is that in small cities, this indicator shows a strong positive correlation (0.84) with *vehicle reuse rate*, a significant association not observed in other city scale groups. A higher cluster coefficient typically implies that the neighbors of a node in the network also tend to connect with each other, thus forming locally dense community structures. In the relatively compact environment of small cities, this local network tightness might make it easier for a vehicle to quickly find the next travel demand in its surrounding area after completing a service, thereby effectively improving the continuous reuse efficiency of vehicles. It is noteworthy that, as shown in Table 5.8, within the analyzed sample of small cities, their *average cluster coefficient* values themselves exhibit greater internal variability compared to other city groups. This larger variability might provide conducive statistical conditions for observing the strong correlation between this indicator and *vehicle reuse rate*, making the effect of the cluster coefficient more significant.

Table 5.9: Descriptive Statistics for Road Density by City Size

| City Size | Count | Mean | Std | Min | Max | Variance |
|---|---|---|---|---|---|---|
| Small | 8.0 | 7.393740 | 4.317884 | 1.593831 | 12.505276 | 18.644122 |
| Medium | 19.0 | 7.660668 | 3.425046 | 3.067729 | 13.961651 | 11.730943 |
| Large | 8.0 | 9.501724 | 1.492357 | 7.225536 | 11.219437 | 2.227128 |

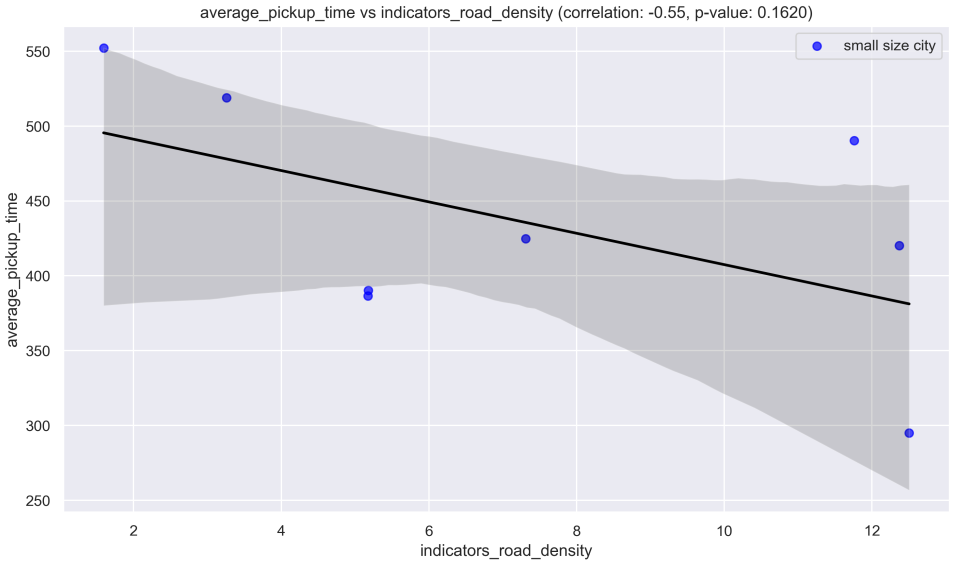



Figure 5.11: Average Pickup Time vs. Road Density in Small Cities

Regarding road density indicators, *road density* and *link density* in medium-sized cities exhibit a relatively significant negative correlation with *average pickup time*. This aligns with the common un-

derstanding that higher density leads to higher accessibility between passengers and vehicles, thereby naturally shortening pickup times. However, in cities of other scales, these density indicators do not show the same correlation. As indicated in Figure 5.9, both *road density* and *population density* demonstrate a very strong positive correlation in medium-sized cities, yet no significant relationship was observed between *population density* and *average pickup time*. This suggests that the road network density, as a structural feature, might play a relatively independent role in medium-sized cities. Meanwhile, Table 5.9 shows that road density in large cities is generally at a high level with relatively small variance. In this scenario, due to the limited range of variation in the independent variable, even if road density impacts pickup time, it might be difficult to detect statistically. Further analysis reveals that a close positive correlation is also observed between *road density* and *population density* in the small city sample. However, neither *road density* nor *population density* showed a significant statistical association with *average pickup time* in small cities. Although a larger range of variation theoretically facilitates the identification of effects, in instances of very small sample sizes, such substantial variation, if accompanied by considerable random noise (Figure 5.11), might conversely impede the reliable detection of any underlying trends. Therefore, based on the current data, higher road network density can be considered a likely positive factor in enhancing the pickup efficiency of SAV systems. However, the generalizability of this conclusion to large or small cities requires careful consideration.

Regarding population density indicators, it is observed that in small cities, *population density* is positively correlated (0.75) with *avg rides per vehicle*. This could be attributed to the fact that in high-population-density small cities, their limited urban scale allows SAVs to more efficiently respond to and connecting rides when local areas have high demand density, thereby significantly increasing the daily service frequency per vehicle and overall operational intensity. In large cities, however, *population density* shows a strong positive correlation with both *peak concurrent vehicles* and *total vehicles*. Since *population density* in large cities has a strong positive correlation with total *population* (0.83), and the direct impact of population on fleet size is also observable, it is difficult to clearly isolate the independent contribution of *population density* to fleet size requirements, separately from the effect of overall population scale. In this context, the *population density* of large cities serves more as a reflection of their massive overall demand volume. Conversely, in the small city group, the correlation between *population density* and *population* is weaker (0.42), and no other indicators were found to have a significant correlation with *average rides per vehicle*. Therefore, the analysis of small cities can better reflect the relatively independent impact of *population density* on matching efficiency. Overall, the impact pattern of *population density* exhibits a significant dependence on city scale. In small cities, it appears to primarily enhance per-vehicle operational efficiency by increasing demand density and enabling faster SAV response times in a limited area. In large cities, its effect is more integrated into the determining role of the city's total population scale on overall fleet size configuration.

Recalling the phenomenon previously observed in the quantitative analysis(Section 5.2.1), where large cities performs lower *average pickup times* and *extra mileage ratio*, it was assumed that this is principally attributed to their higher *population density*. The current analysis of the impact of *population density* across different city scales, combined with the key finding from the all-city sample analysis (see Table 5.3) which established a general negative correlation between *population density* and *average pickup time* (correlation coefficient of -0.5111, P-value $< 0.05$), provides robust evidence to validate this hypothesis. Specifically, the observation from this analysis is that the *population density* of large cities mainly reflects their substantial demand level(population). This considerable and spatially concentrated demand, inherently associated with the high average *population density* characteristic of the large city group (Table 5.7), creates an environment that is highly conducive to efficient vehicle-passenger matching. This explains why the general pattern of higher *population density* leading to shorter *average pickup times* is particularly evident in large cities. The high *population density* in large cities significantly increases the likelihood of SAVs quickly finding nearby passengers, directly leading to shorter *average pickup times* and, consequently, a reduced *extra mileage ratio*.

## 5.3 System Parameter Analysis

This section aims to evaluate the performance of the system under the benchmark parameter settings against variations in other key operational and model parameters. By examining the stability or sensitivity of system performance indicators to parameter changes, this analysis can provide insights into the flexibility boundaries and risk warnings for operational strategy formulation.

### 5.3.1 Service Level Parameter: Impact of Maximum Pickup Delay ($\Delta t^{p,\max}$)

This section analyzes the impact of the $\Delta t^{p,\max}$ tolerable by passengers on system performance. As a important operator-defined service level parameter, it directly constrains the availability of vehicle assignments.

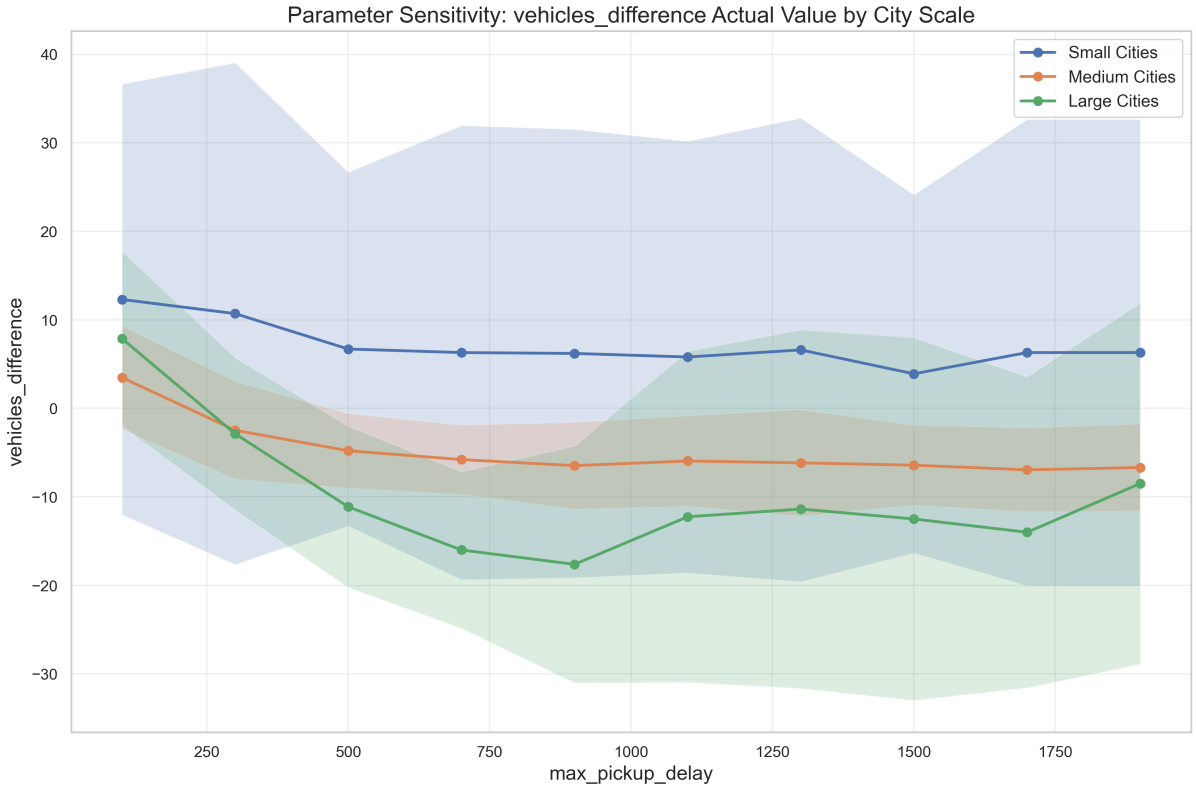



Figure 5.12: Difference in the number of vehicles vs. $\Delta t^{p,\max}$

To begin with, it is needed to determine the impact of this parameter on fleet size. In the simulation setup, the initial fleet size is designed to precisely meet the demand. Figure 5.12 shows that in extreme cases where the $\Delta t^{p,\max}$ is very low, the required fleet size (inferred from the number of additionally triggered vehicles) slightly increases. This is because the conditions for utilizing existing vehicles become excessively strict, requiring the system to frequently trigger new vehicles to satisfy demand. As the $\Delta t^{p,\max}$ increases, the required fleet size tends to stabilize and remains at the level that was previously calibrated, where vehicle resources are relatively sufficient or just meet the demand. This indicates that, within the current simulation environment and initial fleet size settings, once the $\Delta t^{p,\max}$ exceeds a certain threshold, its variation has a minor impact on the final core fleet size required; it primarily acts as a threshold for filtering feasible vehicles.

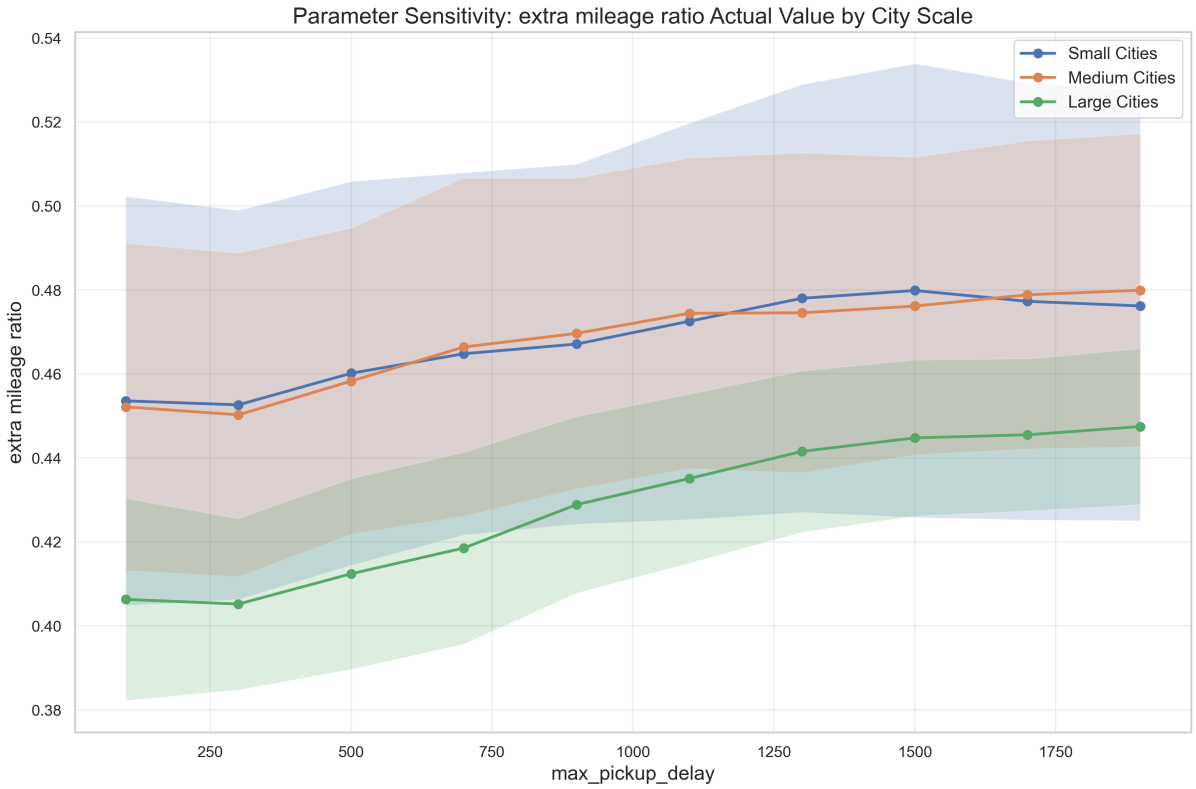
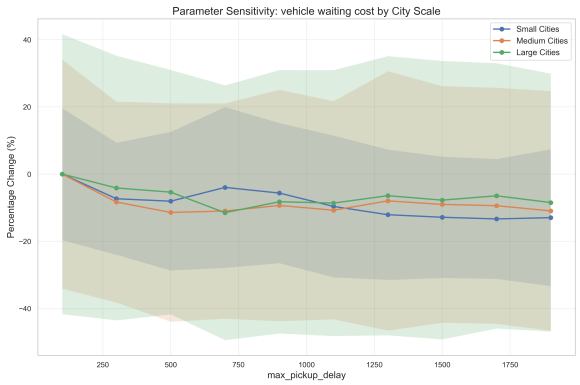
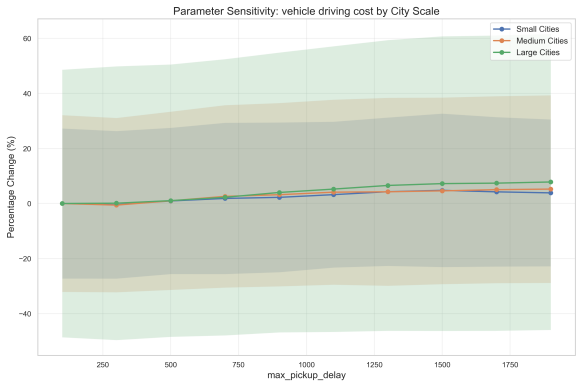



Figure 5.13: Extra Mileage Ratio vs. $\Delta t^{p,\max}$

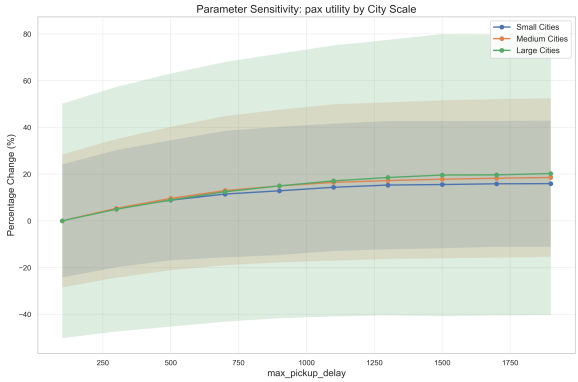
As shown in Figure 5.13, a greater tolerance for $\Delta t^{p,\max}$ leads to an increase in vehicle extra mileage. Since the vehicle's extra mileage is mainly composed of pickup distance, while the service travel mileage for passengers is relatively fixed. This implies that the overall vehicle operational costs will rise as the increase of tolerance for pickup delay, which allows vehicles that were previously unable to serve due to excessive distance or insufficient time to become potential options, providing the system with more opportunities to undertake these long-distance pickup tasks. The potential trade-off of this strategy might be to achieve lower vehicle waiting times or improve certain aspects of passenger experience (although intuitively, passenger waiting times would increase). Further analysis of the changes in relevant cost indicators is conducted to investigate this trade-off.




(a) Change in vehicle waiting cost




(b) Change in vehicle driving cost




(c) Change in passenger cost

Figure 5.14: Impact of different pickup delay tolerances on various cost indicators (percentage change)

As shown in Figure 5.14, increasing the pickup delay tolerance leads to an increase in both passenger costs and vehicle driving costs, while vehicle waiting costs decrease accordingly. This reveals a clear trade-off: the system accepts longer pickup distances and longer passenger waiting times in exchange for reduced vehicle waiting time. When vehicles are allowed more time to reach passengers, they have a higher probability of formulating routes that eliminate waiting time at the pickup location, thereby avoiding associated costs. Based on this observation, the following strategic recommendations can be offered to operators: If the primary goal is to control the certain driving component of operating costs (e.g., in scenarios sensitive to high fuel/electricity prices or vehicle maintenance costs) and simultaneously optimize the passenger waiting experience, setting a relatively lower $\Delta t^{p,\max}$ is likely preferable. Conversely, if reducing vehicle waiting time is the main priority due to specific factors (such as high parking fees or a strong emphasis on high vehicle utilization), increasing the $\Delta t^{p,\max}$ could be considered. However, it must be noticed that this typically comes at the expense of driving efficiency and passenger waiting time. Operators need to evaluate whether the reduction in waiting costs sufficiently compensates for the increased driving costs and potential negative impacts on passenger satisfaction.

### 5.3.2 Joint Sensitivity Analysis of Willingness-to-Share Resistance Factor ($\omega$) and Shared Discount ($\delta$)

This section utilizes the *shared passengers ratio* as the primary KPI. This metric is selected because it directly measures the outcome of the ride generation stage, reflecting the theoretical pooling potential determined by passengers' preferences on pooling ($\omega$) and the offered fare discount ($\delta$). This distinguishes it from the *pooling ratio*, which measures the final realized share rate after the subsequent vehicle assignment and optimization phase. The difference between these two metrics, is further analyzed in Appendix I. Given that $\omega$ and the fare discount offered for shared rides jointly determine the attractiveness of pooling rides and directly influence the *shared passengers ratio* at the ride generation stage, this study conducts a joint sensitivity analysis of these two parameters. Their interaction effects on system performance, particularly on pooling behavior and related efficiency metrics, are systematically evaluated across different city scales (represented by Purmerend, Haarlem, and Rotterdam as small, medium, and

large cities, respectively) using a grid search methodology. Four complementary plot types are utilized to analyze the interaction effects comprehensively, including trend plots, interaction effect plots, gradient heatmaps, and heatmaps. Because the figures are complementary and require cross-referencing for a comprehensive analysis, this section will first describe the observed phenomena from each plot type, and then proceed with a detailed, integrated analysis.

The first type of plot to analyze is the trend plots for the three cities. The left and right subplots in each figure respectively illustrate the impact of $\delta$ on the shared ratio when $\omega$ is fixed, and the impact of $\omega$ on the shared ratio when $\delta$ is fixed.

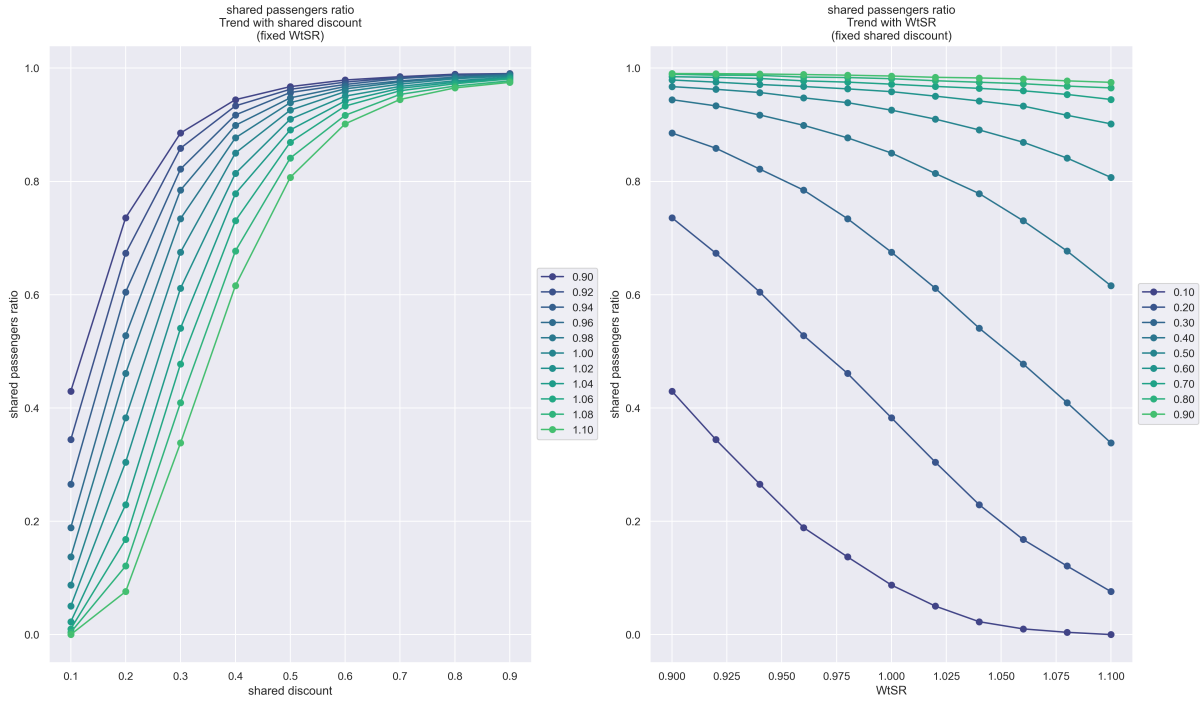



Figure 5.15: Trend Analysis of $\omega$ and $\delta$ in Rotterdam

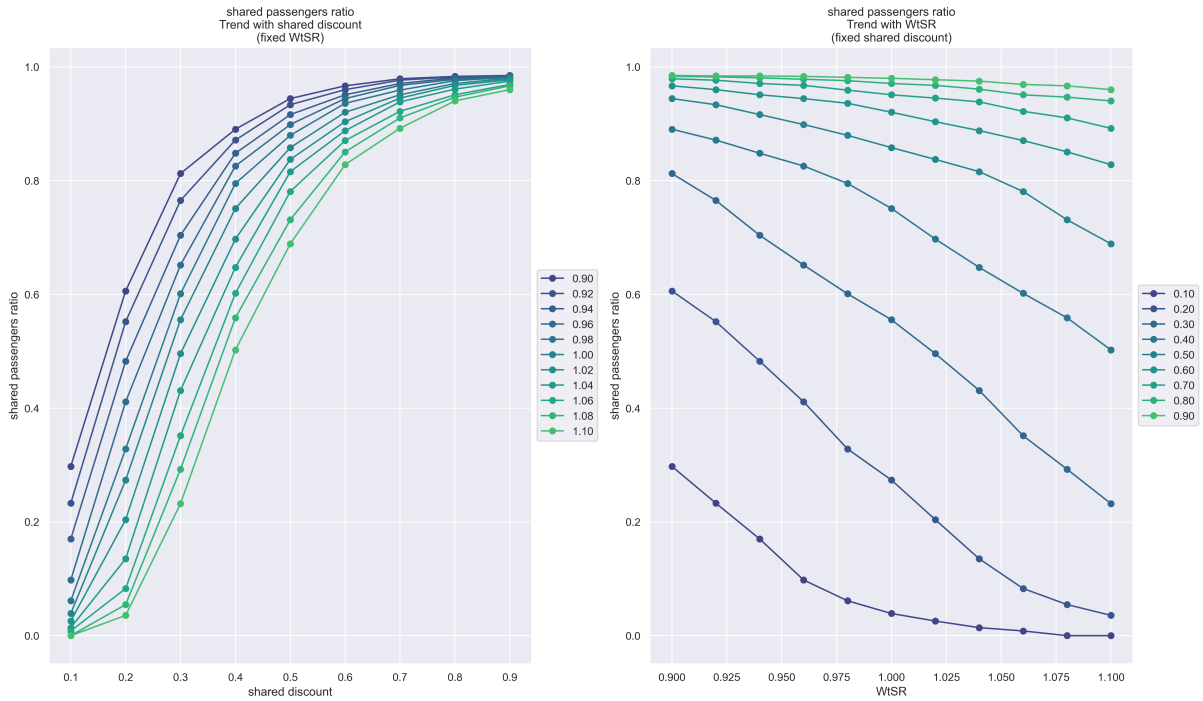



Figure 5.16: Trend Analysis of $\omega$ and $\delta$ in Haarlem

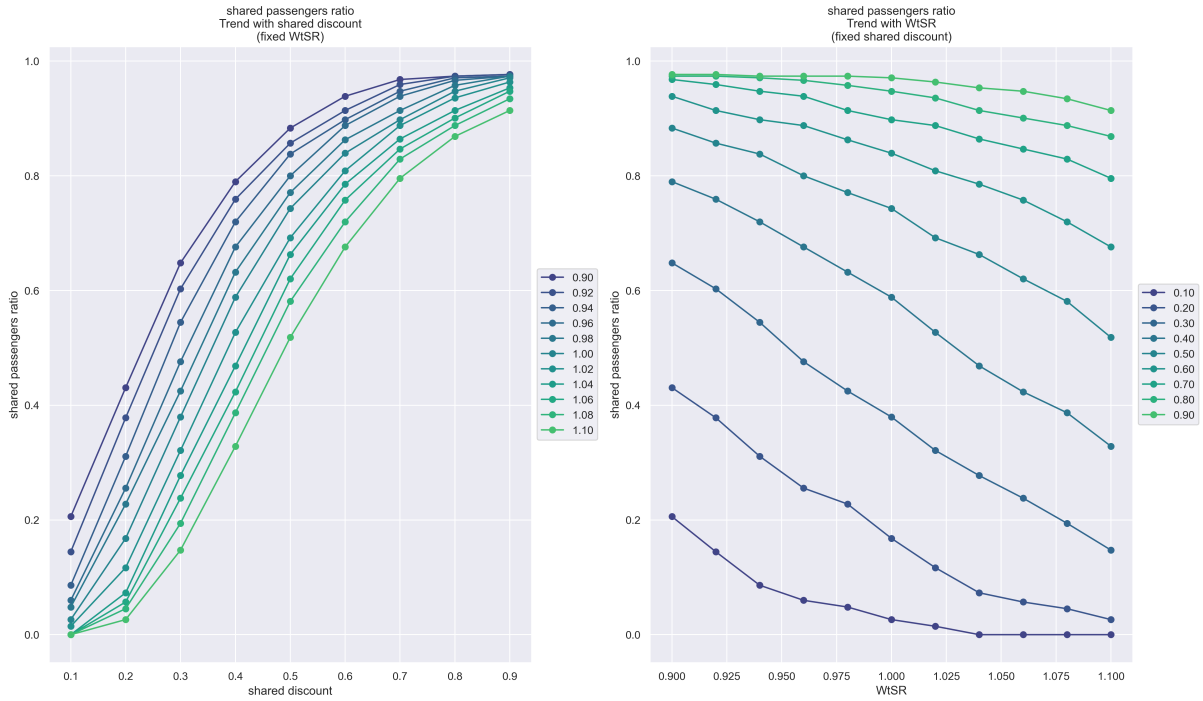



Figure 5.17: Trend Analysis of $\omega$ and $\delta$ in Purmerend

Figures 5.15, 5.16, and 5.17 illustrate the changing trends of *shared passengers ratio* under various parameter combinations. It can be observed in the left subplots in each figure (fixed $\omega$) that, as $\delta$ increases, the *shared passengers ratio* in large cities tends to saturate (approaching 100%) more rapidly than in small and medium-sized cities. Concurrently, even at very low discount levels, the *shared passengers ratio* in Rotterdam is considerably higher than in the other two cities. It can also be observed that with a fixed $\omega$, the growth rate of the *shared passengers ratio* changes significantly around a $\delta$ of 0.2. For lower $\omega$ values, the rate of increase becomes faster, while for higher $\omega$ values, the rate of increase slows down. This suggests that passengers are relatively price-sensitive in the 0.1-0.3 discount range. When $\omega$ is lower (indicating a greater acceptance of the inconveniences associated with sharing), the growth rate is higher in the 0.2-0.3 discount range. Conversely, for passengers with lower $\omega$, the growth rate is higher in the 0.1-0.2 discount range. This phenomenon is particularly evident in medium and large cities, which means that for these cities, the *shared passengers ratio* is more sensitive to changes in the 0.1-0.3 discount range.

The right subplot in each figure (fixed $\delta$, varying $\omega$) provides further insights into the moderation effect of city scale. It can be observed that when $\delta$ is within a high range, the *shared passengers ratio* curve for large cities appears relatively flat; even as $\omega$ increases from lower to higher values, the decrease in the *shared passengers ratio* is limited. However, when $\delta$ is within a low range, the *shared passengers ratio* curve for large cities declines more steeply with an increase in $\omega$ compared to small cities. At first glance, this might suggest that, under low discount policy, users in large cities are more sensitive to the changes in $\omega$. Nevertheless, a crucial context must not be considered: even at the lowest discount levels, the initial *shared passengers ratio* in large cities is significantly higher than small cities. This implies that although the rate of decline in the *shared passengers ratio* is rapid with increasing $\omega$ under low discounts, given its higher starting base, an acceptable *shared passengers ratio* can still be achieved with this discount level.

Clear trends and initial judgments have been identified from the trend plots. However, precisely and objectively assessing the combined effect of the two parameters and the system's sensitivity across the entire parameter space is challenging with these multi-series line charts alone. To overcome this limitation, the following analysis introduces interaction effect heatmaps and gradient heatmaps, to serve as a complement to the trend plots, quantifying the patterns observed in the trend plots with more specific numerical values.

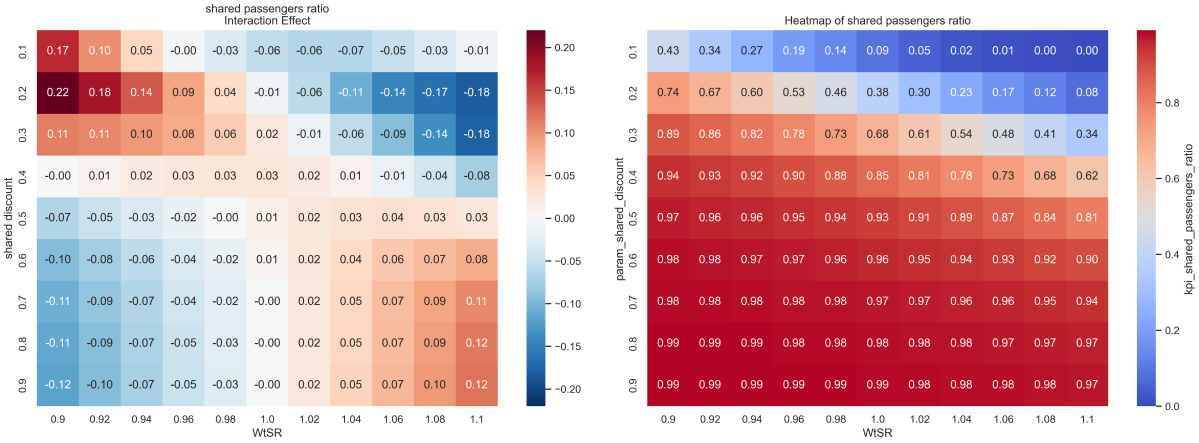



Figure 5.18: Interaction Effect Analysis of $\omega$ and $\delta$ in Rotterdam

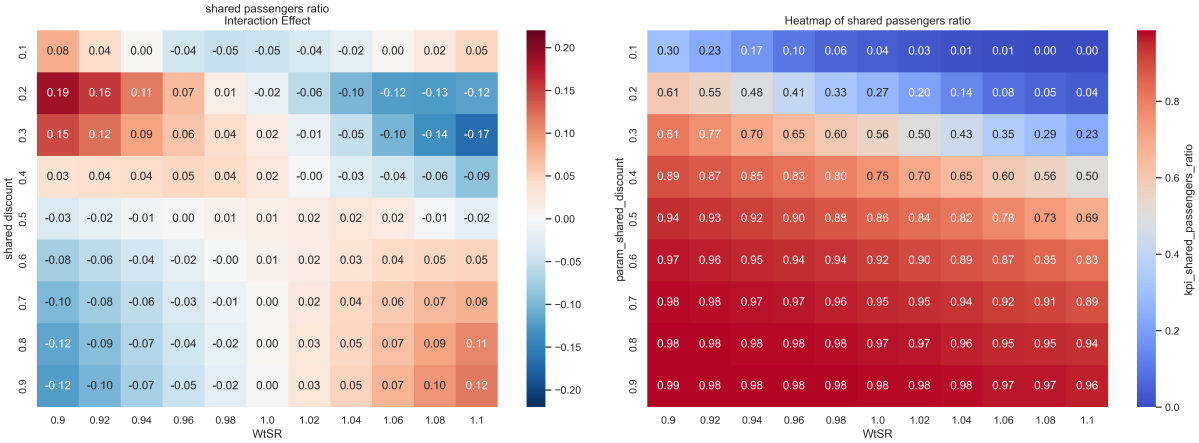



Figure 5.19: Interaction Effect Analysis of $\omega$ and $\delta$ in Haarlem

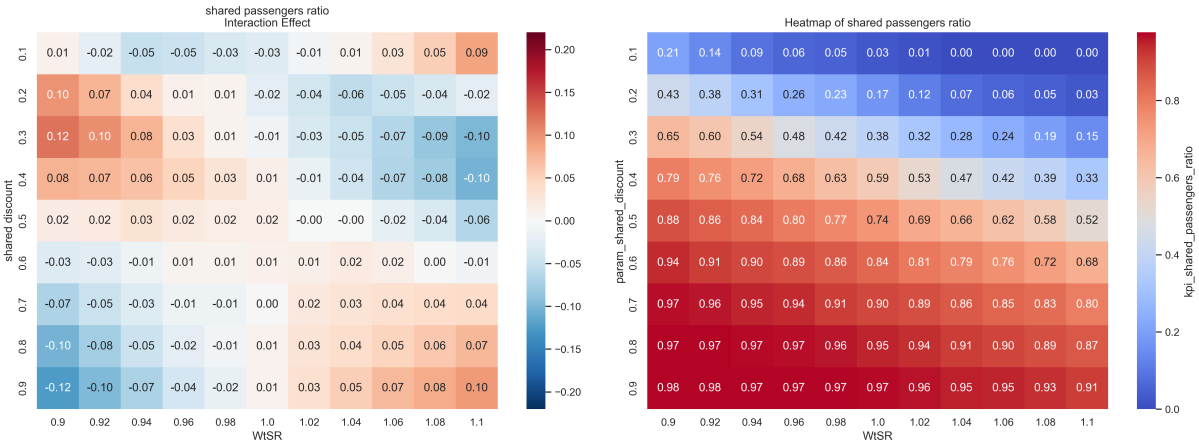



Figure 5.20: Interaction Effect Analysis of $\omega$ and $\delta$ in Purmerend

Figure 5.18, 5.19, and 5.20 present the interaction effect heatmaps for the two parameters (left subplot) and the heatmaps for the *shared passengers ratio* (right subplot).

The interaction effect heatmap quantifies the extent to which the joint impact of two input parameters (e.g., $\omega$ and $\delta$) on an outcome metric (e.g., *shared passengers ratio*) deviates from the sum of their individual average effects.

Simply put, the sign of the interaction effect measures whether the parameter combination results in a combined positive effect (where the outcome exceeds the sum of individual effects) or an offsetting effect (where the outcome is less than the sum of individual effects), while its absolute value indicates the strength of this interaction.

The formula for its calculation is defined as follows:

$$\text{Interaction Effect}_{ij} = \text{Observed Value}_{ij} - \text{Row Mean}_i - \text{Column Mean}_j + \text{Grand Mean}$$

Where:

- Observed Value$_{ij}$ is the actual observed value of the outcome metric for the parameter combination $(i, j)$.
- Row Mean$_i$ refers to the average value of the outcome metric across all levels of $j$ when the $i$-th parameter (e.g., $\delta$ fixed at its $i$-th level) is held constant. This represents the average main effect of the row parameter.
- Column Mean$_j$ refers to the average value of the outcome metric across all levels of $i$ when the $j$-th parameter (e.g., $\omega$ fixed at its $j$-th level) is held constant. This represents the average main effect of the column parameter.
- Grand Mean is the overall average of the observed values across all parameter combinations.

The resulting Interaction Effect$_{ij}$ value represents the deviation of the parameter combination $(i, j)$ from the expected value based on the simple sum of its independent main effects. A positive value indicates a synergistic effect, meaning the combined effect is greater than the sum of the individual effects, while a negative value indicates an antagonistic or saturation effect, where the combined effect is less than the sum of the individual effects.

The interaction effect heatmap can be viewed as a numerical representation of the trend plots. For example, as observed in the trend plots for Rotterdam, at a $\delta$ of 0.2, the slope of the *shared passengers ratio* increase becomes smaller for individuals with low $\omega$, whereas it becomes larger for those with high $\omega$. This is reflected in the interaction effect heatmap as: at a $\delta$ of 0.2, the interaction effect is negative for low $\omega$ individuals and positive for high $\omega$ individuals. This indicates that, at this discount level, the rate of increase in the *shared passengers ratio* as the discount increases is below and above the average, respectively. Using the interaction effect heatmap, a more intuitively comparasion can be obtained for assessing the differences in the impact of parameter combinations on the *shared passengers ratio* across cities of different scales. Simultaneously, it is necessary to consider the heatmap of the *shared passengers ratio* itself to obtain information about the actual absolute values.

Regarding cities of different scales, some clear distinctions can be observed. First, there are significant differences in the strength of the interaction effects among the three cities. Visually, the overall strength of interaction effects tends to be higher in large cities and lower in small cities. Specifically, in Rotterdam, the intervals with higher absolute values of positive and negative interaction effects occur at lower discount rates (0.1-0.2 for Rotterdam, compared to 0.2-0.3 for the other two cities). Its interaction effect strength is higher compared to the other cities, and the distribution range of parameter combinations exhibiting strong interaction effects is also more concentrated. For small cities, however, significant interaction effects appear in both high and low discount ranges. It is noteworthy that small cities exhibit a relatively significant positive interaction effect at the parameter combination of $\delta = 0.1$ and $\omega = 1.1$, a trend contrary to that observed in large cities. This implies that for individuals in small cities with a higher $\omega$, the actual *shared passengers ratio* in the low discount range exceeds the average expectation. However, the actual value of the *shared passengers ratio* is 0, so the positive interaction effect means that this outcome is actually higher than the value predicted by simply adding the average effect of that row ($\delta$ 0.1) and that column ($\omega$ 1.1). This indirectly reflects the difficulty in stimulating sharing among passengers unwilling to share in small cities within the low discount range. The positive interaction effect here can be explained by the fact that the sharing ratio cannot be less than zero, so when the predicted result from adding independent effects approaches or falls below zero, the actually observed zero value generates a positive interaction effect value.

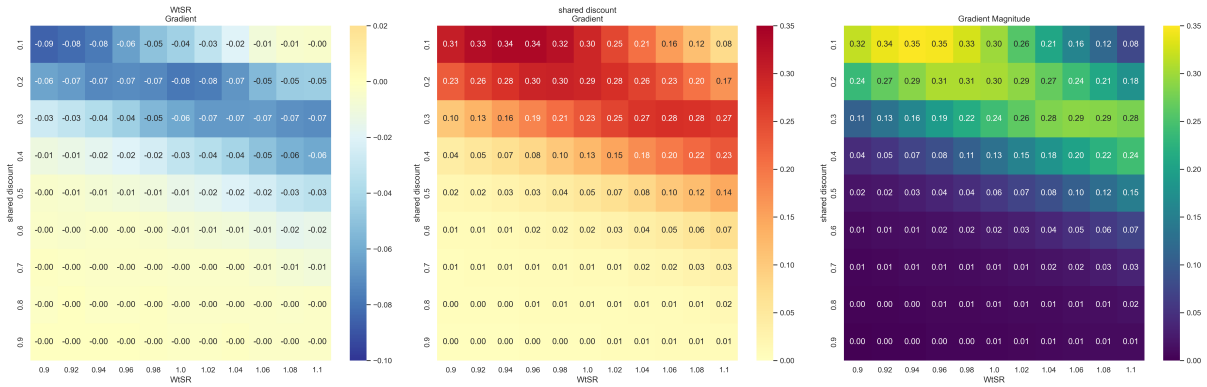



Figure 5.21: Gradient Heatmap Analysis of $\omega$ and $\delta$ in Rotterdam

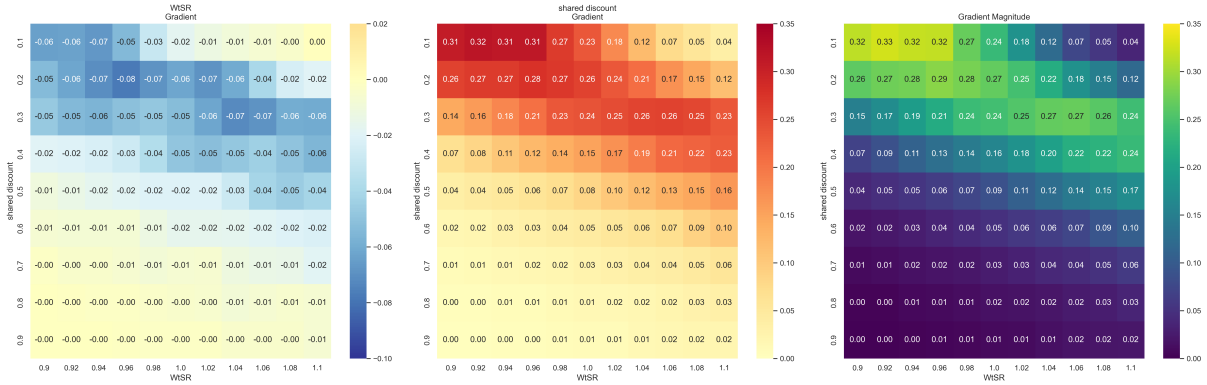



Figure 5.22: Gradient Heatmap Analysis of $\omega$ and $\delta$ in Haarlem

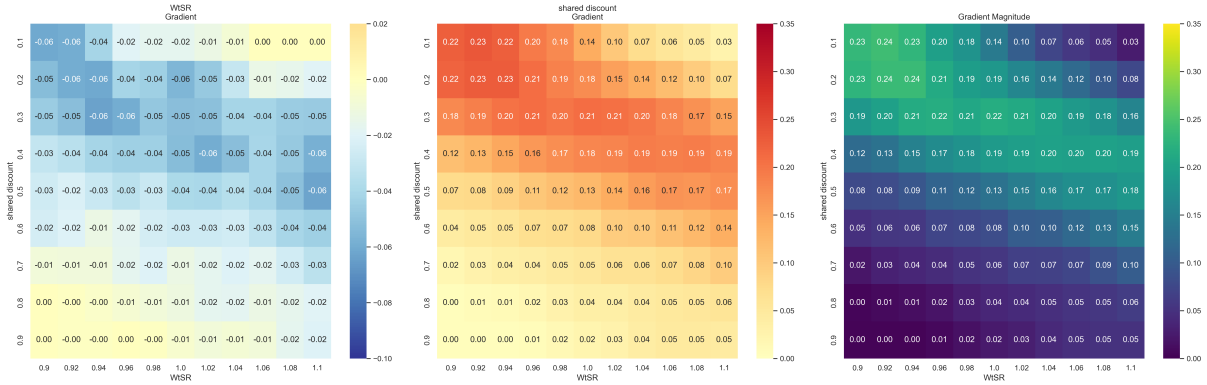



Figure 5.23: Gradient Heatmap Analysis of $\omega$ and $\delta$ in Purmerend

Figure 5.21, 5.22, and 5.23 display the gradient heatmaps for the three cities. These gradients are calculated using the *numpy.gradient* function(Harris et al., 2020). Specifically, the first subplot represents the gradient with respect to $\omega$ (horizontal direction), while the second subplot represents the gradient with respect to $\delta$ (vertical direction). The *numpy.gradient* function automatically selects neighboring points for each parameter combination to compute the gradient in that direction at the current parameter setting. For instance, the gradient maps for $\omega$ consistently show negative values, indicating that the *shared passengers ratio* decreases as $\omega$ increases, with the absolute value representing the rate of decrease. The third subplot is the gradient magnitude heatmap, where the value for each parameter combination is the Euclidean norm of the gradients in the two parameter directions. A higher value in this heatmap signifies greater sensitivity of the *shared passengers ratio* to both parameters at that specific combination.

Similar to the interaction effect maps, the gradient heatmaps reveal that the sensitivity strength to different parameter combinations is markedly higher in large cities compared to small cities. Further-

more, for large cities, the high-sensitivity ranges are more concentrated, whereas the sensitivity range in small cities is more dispersed. Moreover, an axis of symmetry running from the top-left to the mid-right is observable in the third subplot for all three cities, representing the parameter ranges where the city exhibits the highest sensitivity. Compared to the sensitive ranges in other cities, the sensitive ranges in large cities is generally shifted towards lower discount ranges, meaning the axis of symmetry is positioned higher overall. This indicates that passengers in large cities, regardless of changes in their $\omega$, are consistently more sensitive to lower discount ranges.

After separately presenting the findings from the interaction effect heatmaps and gradient heatmaps, they can now be analyzed in a combined view to determine the most reasonable parameter combinations for different cities. In other words, the question lies in how to design pricing policies for passengers with varying willingness to share, considering different city scales? As previously explained, the gradient heatmaps reveal distinct axes of symmetry; the darker the color along these axes, the more sensitive passengers are under those parameter combinations. However, knowing the sensitivity alone is insufficient, as high sensitivity might lead to a greater increase in the *shared passengers ratio*, while low sensitivity might result in a more stable ratio. Sensitivity itself does not rerelcts the strategical desirability of a parameter combination. Therefore, it is necessary to combine the interaction effect heatmaps and the heatmaps of the *shared passengers ratio* to identify the optimal combination of the two parameters. The analysis aims to identify parameter combinations that exhibit synergistic effects in high-sensitivity regions and combinations that offer stable, high sharing ratios in low-sensitivity regions. Conversely, combinations with antagonism effects in high-sensitivity areas should be avoided. Such combinations imply inefficient discount settings that fail to achieve the expected increase in sharing ratio. Furthermore, due to the high sensitivity associated with these combinations, even minor unfavorable changes (e.g., increased reluctance to share or a slightly lower discount) could lead to a significant drop in the sharing ratio. Taking Rotterdam as an example, by observing Figure 5.18, the strongest positive interaction effect is found occuring at a $\delta$ of 0.2 and a $\omega$ of 0.9. This indicates that under this parameter combination, the increase in the sharing ratio significantly exceeds expectations. From Figure 5.21, the gradient heatmap indicates high sensitivity for the same parameter combination. Thus, for individuals already inclined to share (i.e., those with low $\omega$), setting the discount at 0.2 can more effectively increase their sharing rate compared to other combinations. An opposing example arises when the discount ratio is 0.3 for individuals with high $\omega$. Here, a significant negative interaction effect is observed, and the gradient heatmap indicates high sensitivity. So it can be concluded that this parameter combination is inefficient and unstable for increasing the sharing ratio. The inefficiency is evident from the negative interaction effect, and the instability stems from the high sensitivity. The gradient heatmap in $\omega$ direction shows that in this region, an increase in $\omega$ (meaning individuals become less willing to share) results in a larger decrease in the sharing ratio. Therefore, this region should be avoided as much as possible in favor of more stable areas. For instance, at a discount rate of 0.5, there is a slight positive interaction effect, and passenger sensitivity is low. Observing the KPI heatmap reveals a sharing ratio as high as 60% at this point. The conclusion is that if engaging this segment of passengers is necessary, a 0.5 discount rate represents a reasonably effective choice. Additionally, considering the axis of symmetry from the gradient heatmap and the sharing ratio heatmap together, the goal should be to make passengers within the high-sensitivity region (along the axis) towards the direction of higher sharing rates, specifically towards the bottom-left corner. This area offers not only high sharing rates but also lower sensitivity; even with negative interaction effects, a considerable sharing ratio can still be achieved. Therefore, certain optimal and undesirable parameter combinations can be identified for Rotterdam. In summary, for passengers with a high willingness to share(Low $\omega$), a 0.2 discount rate yields the most efficient increase in sharing ratio. For those with low willingness to share(high $\omega$), a 0.3 discount rate should be avoided; instead, raising it to 0.5 or higher is preferable, although this segment might offer limited profitability for the operator. Applying the same analysis method to Haarlem and Purmerend reveals differences primarily concerning passengers with low willingness to share. For these cities, discount rates need to be increased to 0.6 and 0.7, respectively, to achieve similar effects. However, for passengers with high willingness to share, the optimal discount rate remains 0.2 for Haarlem, while it is 0.3 for Purmerend.

Furthermore, examination of the heatmap analysis results within this theoretical framework reveals a significant and consistent system behavior pattern across different city scales. Cities with higher endogenous ride-pooling potential and requiring higher calibrated $\omega$ values to fit real-world data, such as large cities like Rotterdam, show much higher sensitivity and stronger interaction effects when parameters change. The combination of high sensitivity and strong interaction effects means that, on the one hand, in certain parameter regions (such as low discounts and low $\omega$), strong positive joint effects indicate substantial opportunities to efficiently improve sharing ratio through targeted strategies like pricing. On

the other hand, in other regions (such as medium discounts and high $\omega$), the combination of strong negative joint effects and high sensitivity results in a significant operational risk: the system not only operates inefficiently, but is also extremely sensitive to minor parameter changes or external disturbances, which can easily lead to a rapid decline in performance. The high calibrated $\omega$ value itself already signals the existence of restraining factors. Therefore, effective operation in large cities requires highly refined management, with precise identification and exploitation of opportunity zones, as well as careful avoidance of high-risk zones, and a high degree of sensitivity in parameter tuning. The heatmap of *shared passengers ratio* thus provides a basis for identifying these different characteristic regions.

### 5.3.3 Scale Factor Analysis ($s$)

In previous sections, due to computational constraints, all analyses of demand were based on 1% of the real travel volume. However, it is unknown that how would various indicators change if the demand scale were increased. This forms the primary motivation for the sensitivity analysis of $s$ in this section. This chapter will explore how the absolute values of KPIs, derived from small demand samples, change by adjusting $s$, which represents the proportion of simulated demand relative to the total real travel volume. Such analysis aims to provide insights into the expected service levels or fleet efficiency under varying demand volumes or SAV penetration rates, offering valuable guidance for operators in predicting performance across different scenarios.

Given that this subsection focuses more on the impact of demand scale rather than the detailed optimization of initial resource allocation, a fixed initial fleet size parameter ($p_{init}$set at 15% of the demand volume) was uniformly adopted in this series of $s$ sensitivity analysis. This setting ensures that all simulations have sufficient vehicle resources while excluding the variable impact introduced by additional vehicles, and it is regarded as an acceptable implementation approach at the current research stage.

In this section, Delft (as a representative of medium-sized cities with a lower population) and Purmerend (as a representative of small-sized cities) are selected for case analysis. Although these two cities belong to different city scale groups, their population sizes do not exhibit significant differences(109577 and 95168 respectively). Due to computational constraints in $\omega$ calibration process, this section could not conduct simulation analysis for larger-scale cities. Therefore, it remains uncertain whether medium-sized cities like Haarlem or large-sized cities like Amsterdam would have the same patterns as the two cities selected. Nevertheless, this section has identified some patterns in cities with lower population sizes that may hold general applicability. The experimental process strictly followed the steps below as the demand generation process to ensure systematic and comparable analysis. First, the theoretical total daily travel demand for each city was calculated based on its population base and average travel frequency. Subsequently, for a series of predefined $s$ values (ranging from 0.01 to 0.10 with a step size of 0.01), simulated travel demand datasets of corresponding scales with spatiotemporal distribution characteristics were generated for each city.

To ensure that the simulation results closely reflect the actual ride-sharing tendencies of each city, $\omega$ calibration was conducted separately for each $s$ in both cities. The calibration objective was to make the simulated ride-sharing rate as close as possible to the actual observed ride-sharing rate extracted from the ODiN dataset for each city. The specific calibration method is detailed in Section 5.1.1. The following results will present and analyze the trends of these KPIs with varying $s$ values, and a comparative discussion of the two representative cities will be provided.

Table 5.10: Regression trends of key KPIs with $s$ in *Purmerend* and *Delft*

| KPI | *Purmerend* | | *Delft* | |
|---|---|---|---|---|
| | Slope | $R^2$ | Slope | $R^2$ |
| Total Fleet Size | 6320.0 | 0.98 | 5269.1 | 0.96 |
| Avg. Vehicle Waiting Time | -1776.0 | 0.68 | -1637.3 | 0.84 |
| Extra Mileage Ratio | -1.020 | 0.91 | -0.839 | 0.70 |

From the perspective of fleet size requirements, total fleet size increases almost linearly with the rise in $s$. A linear regression analysis shows that, in *Purmerend*, total fleet size increases with a slope of 6320.0 per unit increase in $s$ ($R^2 = 0.98$), while in *Delft*, the slope is 5269.1 ($R^2 = 0.96$), indicating a strong linear relationship between demand scale and fleet size in both cities. This trend aligns with previous results based on the quantitative correlation analysis across different city scales, indicating that the total number of required vehicles is directly related to the total demand volume. Interestingly, small

cities exhibit a growth trend similar to that of larger cities when the demand scale undergoes significant changes (recall Section 5.2.3, where the population of small cities did not show a strong correlation with fleet size), which might indicate that the variance and absolute value of population of small and medium-sized cities are not significant enough within their group. More notably, the demand growth patterns in cities of both scales show a certain similar linear trend. This suggests that the impact of demand scale on fleet size has a level of general applicability, providing a basis for inferring fleet size trends under larger demand scales from smaller demand samples.

In terms of fleet operational efficiency, multiple indicators collectively demonstrate the impact of demand scale. As $s$ increases, the average vehicle waiting time generally trends downward, with regression analysis showing negative slopes in both cities (*Purmerend*: $-1776.0$, $R^2 = 0.68$; *Delft*: $-1637.3$, $R^2 = 0.84$). At the same time, the extra mileage ratio also exhibits a clear decreasing trend (*Purmerend*: $-1.020$, $R^2 = 0.91$; *Delft*: $-0.839$, $R^2 = 0.70$), indicating a significant enhancement in vehicle utilization efficiency as demand increases. The expansion of demand scale leads to the deployment of more vehicle resources, which reduces non-passenger mileage and idle waiting periods through improved matching efficiency.

In terms of average passenger waiting time, the results show that as $s$ increases, the values remain stable within a specific range. Specifically, in *Delft*, the average passenger waiting time fluctuates between 54.9 and 80.4 seconds, while in *Purmerend*, it ranges from 53.4 to 70.9 seconds. This relative insensitivity to the absolute demand volume suggests that the passenger service level is more likely constrained by other fundamental factors. These factors could include the inherent geographical scale of the cities, which sets a physical limit on pickup travel times, as well as the predefined weight factor$\alpha$ parameters within the model that guide the optimization.

The second consideration in this section's analysis focuses on whether the contribution and behavioral patterns of passengers' $\omega$ to actual ride-sharing outcomes change under different overall demand scale contexts. Understanding this aspect is crucial for assessing how the potential for ride-sharing, which depends on passengers' willingness to share, evolves as the market transitions from an early stage to a mature stage. Additionally, it helps determine whether the passengers' own willingness or simply sufficient demand density play a dominant role in influencing ride-sharing rates at different stages.

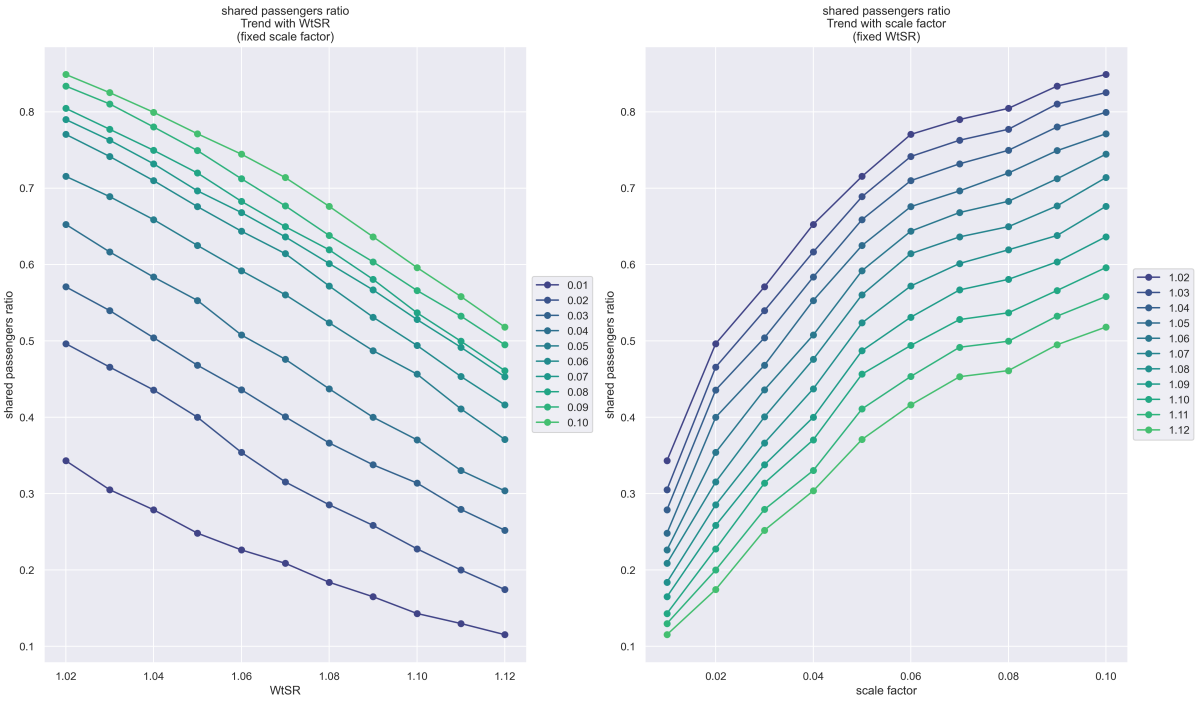



Figure 5.24: Trend Relationship of Shared Passengers Ratio with $\omega$ and $s$ in Purmerend

To more specifically assess the impact of varying demand levels on ride-sharing rates, it is necessary to analyze the trend plot reflecting the influence of $s$ and passengers' $\omega$ on actual ride-sharing rates. At this stage, data from Purmerend is selected for analysis. As shown in Figure 5.24, the left subplot illustrates the trend of ride-sharing rates with varying $\omega$ under different fixed $s$ values. A key observation is that,

assuming a target ride-sharing rate a level of 0.51, which corresponds to the previously calibrated real ride-sharing rate, and examining $\omega$ values required to achieve this target rate under different $s$ levels, it can be noted that as $s$ gradually increases from 0.01 to 0.10, the horizontal spacing between these $\omega$ value points, which reflects the range of variation in $\omega$, shows a diminishing trend. This phenomenon implies that once the demand scale reaches a certain level, the system's sensitivity to $\omega$ decreases. Alternatively, it can be understood that to maintain a relatively stable ride-sharing rate under high demand density, the need for significant adjustments to $\omega$ (choosing different target passengers) is reduced. This trend of scale effect's saturation suggests that at higher demand scales, the response of ride-sharing rates to $\omega$ tends to stabilize. The right subplot reveals the impact of demand scale from another dimension, showing the trend of ride-sharing rates with varying $s$ under different fixed $\omega$ levels. Observing the curves representing different $\omega$ levels, the vertical spacing between them reflects the differences in ride-sharing rates due to varying $\omega$, and this spacing increases as $s$ rises. This phenomenon indicates that although an increase in demand density generally enhances sharing opportunities, at higher demand scales, passengers' own $\omega$ plays a more significant distinguishing role in determining the extent of sharing that can ultimately be achieved. In other words, when potential matching opportunities are abundant, user preferences become the primary constraining factor. This observation suggests that in larger-scale scenarios, the enhanced distinguishing effect of $\omega$ makes targeted operations for user groups with different $\omega$ characteristics, such as applying strategies combining $\omega$ with shared discounts discussed as metioned in previous sections, become increasingly critical.

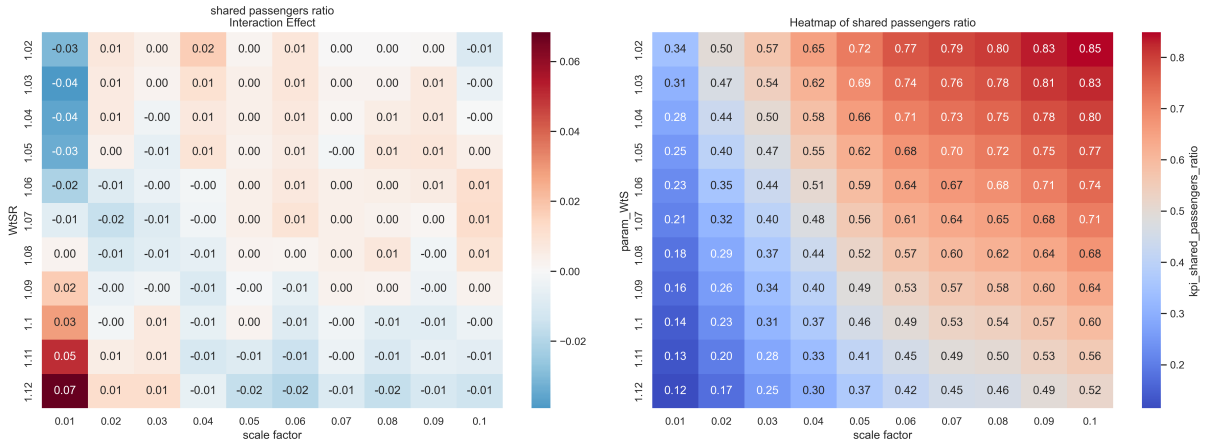



Figure 5.25: Interaction Effect and Heatmap of Ride-Sharing Rate with $\omega$ and $s$ in Purmerend

Figure 5.25 presents the interaction effect heatmap and the ride-sharing rate heatmap for $s$ and $\omega$ on ride-sharing rates. From the interaction effect heatmap, a phenomenon that provides significant operational value can be observed, that is, only when $s$ is at an extremely low level ( 0.01, corresponding to the leftmost column of the heatmap), a notable interaction effect between $\omega$ and $s$ exsists. Specifically, in the higher $\omega$ range, a positive interaction effect is evident, indicating that the actual ride-sharing rate for this group exceeds the level expected from the independent additive effects of high $\omega$ and low demand. Conversely, in the lower $\omega$ range, a negative interaction effect is observed. For operators in the early market stage, this finding suggests that when SAVs' penetration rate is low, focusing on user groups with lower willingness to share might be more beneficial, as they could exhibit ride-sharing rates that exceed anticipated levels at this stage.

However, once $s$ exceeds 0.01, the values in the interaction effect heatmap generally drop sharply and approach zero. This shift reveals that when the demand scale reaches a certain level, the combined influence pattern of $\omega$ and $s$ on ride-sharing rates stabilizes. Specifically, their individual contributions to ride-sharing rates become more independent and predictable. This implies that although $\omega$ remains a critical factor in distinguishing user sharing behavior, and its importance may even grow with increasing demand scale, the sensitivity of strategy outcomes to further changes in overall demand scale decreases when designing and adjusting strategies for different $\omega$ groups. Consequently, a more universal operation strategy can be developed to manage targeted user segmentation based on $\omega$ and corresponding strategies, effectively addressing variations in overall demand scale and thereby enhancing the robustness of operational strategies and the accuracy of system behavior predictions. Combined with the significant interaction effect observed at a $s$ of 0.01, this suggests that in the initial stage of extremely low market

penetration, highly contextual and dynamically adjustable strategies may be necessary for passenger groups with varying willingness to share. Once the system moves beyond this complex interaction phase into a stage where demand effects and $\omega$ effects are relatively independent, although the need for targeted operations based on $\omega$ (such as differentiated pricing or services) remains critical (and may even become more so due to the enhanced distinguishing power of $\omega$), the formulation and application of these strategies can occur within a more stable and predictable framework, with their effectiveness no longer fluctuating with further variations in demand scale.

Trend plots and interaction effects reveals a saturation effect in system performance as demand scale grows. As $s$ increases from low to high levels, its marginal contribution to facilitating ride-sharing gradually diminishes. This phenomenon suggests that a moderate increase in demand density in the early stages can lead to significant improvements in system performance. However, once the demand scale reaches a higher level, the marginal benefits of further expanding demand scale may decrease. For operators, this indicates that limited resources should be prioritized for markets still in lower demand density ranges, which may correspond to small cities or the initial stages of system operation in reality, to achieve more substantial pooling ratio performance. Based on the above analysis, ride-sharing behaviors of users with varying willingness to share exhibit distinct characteristics under different demand scales. Therefore, operators should adjust their strategic directions accordingly. Transitioning from highly targeted and differentiated strategies in extremely low-demand scenarios to standardized frameworks and broad coverage strategies in higher-demand scenarios, system operations need to evolve in alignment with market development stages. Additionally, the observed saturation effect of scale provides valuable insights for prioritizing resource allocation.

## 5.4 Public Transport Competition

This chapter aims to investigate in depth the complex competitive relationship between the SAV system and existing public transport through a series of systematic simulation experiments. The foucus is on systematic evaluation of how external competitive pressure—primarily in the form of PT's pricing strategy ($\delta_{PT}$)—and the macro-level operating conditions, represented by traffic congestion levels ($v_{avg}$), jointly shape the dynamics of the multi-modal transport system. The analysis is not confined to macroscopic market share shifts but delves into the internal demand structure to explore how competition reshapes user choice behavior across different travel distances and city types. More importantly, this section traces the transmission effects of these external factors on the SAV system's internal operational efficiency (e.g., vehicle waiting time, extra mileage ratio), thereby providing a comprehensive and in-depth assessment of the interaction mechanism between them.

## 5.5 Analysis of Public Transport Discount Effects

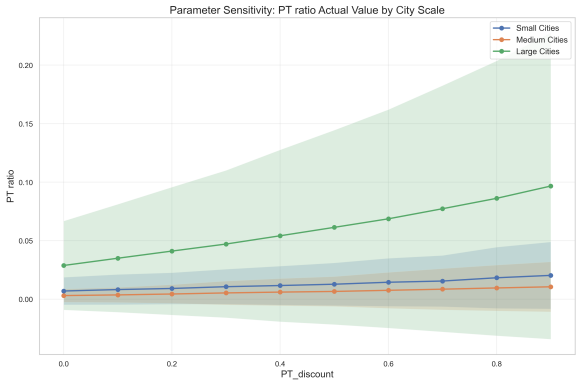
Building upon the previously established baseline scenario, this section systematically adjusts the discount level for public transport to investigate in depth the response mechanisms and operational performance of the SAV system when facing price competition from PT. The results not only reveal the differentiated effectiveness of the PT price lever across different city scales but, more importantly, they show how this external competition reshapes the demand structure of the SAV system and ultimately feeds back into its internal operational efficiency.

Before proceeding with a detailed analysis of the competition structure between SAV and PT, a critical preliminary step is to identify a sample of cities where PT possesses a baseline level of competitiveness. In the initial multi-city simulations, it was notable that the PT mode share (*PT ratio*) in the vast majority of cities remained at extremely low levels. This indicates that in these areas, PT did not pose effective market competition to the SAV. If these cities are included in the analysis, the data would introduce confounding factors and obscure the true competitive mechanisms.

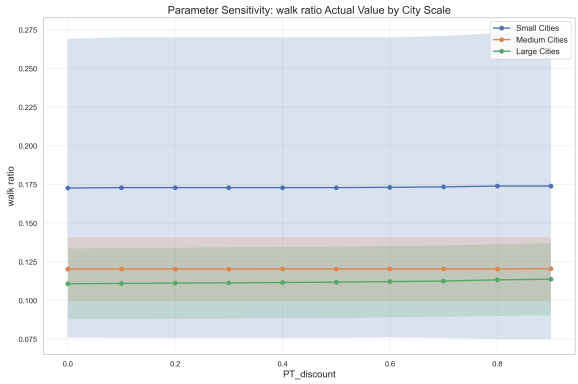
Table 5.11: Cities Constituting the 'Effective Competition Set' (PT Ratio > 1%)

| City | PT Ratio |
|---|---|
| Rotterdam | 19.77% |
| 's-Gravenhage | 15.85% |
| Amsterdam | 11.89% |
| Delft | 4.39% |
| Houten | 4.33% |
| Zoetermeer | 3.75% |
| Nieuwegein | 3.00% |
| Amstelveen | 2.92% |

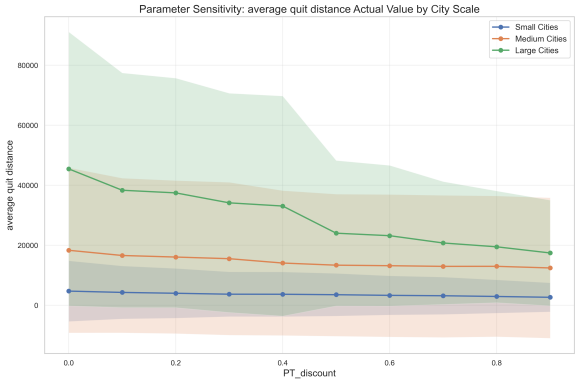
To ensure the validity of the analysis and the reliability of its conclusions, this section establishes a threshold of $PT\ ratio > 0.01$ as the criterion for dividing the cities into two groups. As indicated in Table 5.11, among all simulated cities, a total of 8 are identified as the 'Effective Competition Set'. The remaining cities are classified as the 'Network-Restricted Set', representing those that lack competitiveness against SAVs in the baseline scenario due to their inadequate network infrastructure.




(a) PT mode ratio




(b) Walk filtered ratio




(c) Avg. original travel distance of PT passengers

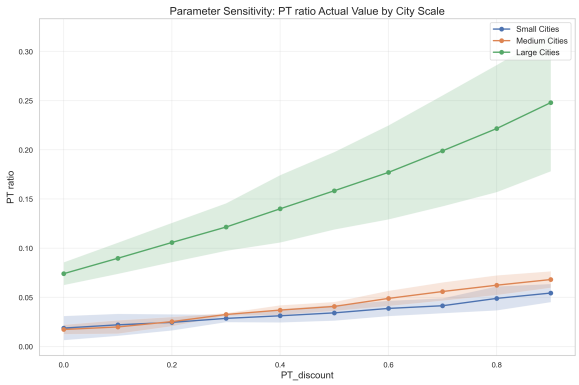
Figure 5.26: Impact of $\delta_{PT}$ on Mode Choice and SAV Operations

Figure 5.26 aggregates the simulation results from all cities, illustrating the overall impact of $\delta_{PT}$ on mode choice. From the perspective of mode share, as shown in Figure 5.26a, the PT mode share exhibits a significant positive correlation with $\delta_{PT}$, which is particularly prominent in large cities. In large cities, where the public transport networks are more developed and thus more competitive, the effect of price adjustments on demand is amplified: the PT mode share increases linearly from approximately 3% at zero discount to about 10% at a 90% discount. In contrast, while the PT mode share in small and medium-sized cities also grows with increasing discounts, its absolute level remains low, indicating that network coverage and connection convenience are more fundamental constraints than price.

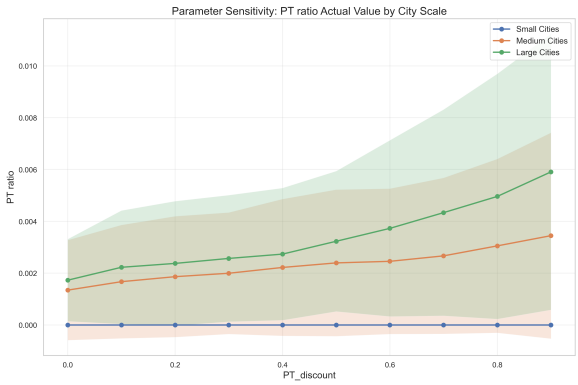
Meanwhile, as illustrated in Figure 5.26b, the share of walking trips is completely unaffected by $\delta_{PT}$

across all cities, holding steady at a constant level. It is noteworthy that in small cities, the absolute value of this share is relatively higher, and the variability among these cities is substantial. This observation, combined with their extremely low PT mode share, further suggests that in small cities, the direct competitive relationship between SAV penetration and the rail-dominated PT system in this study is weak. The constancy of the walking share is a direct consequence of the model's design: the walking mode in this study is not modeled as a direct competitor to SAV or PT. Instead, it is designed as a filter to screen out unreasonable SAV or PT trips. The constant value of the walking share precisely reflects the fixed proportion of such trips within the total demand. This methodological approach ensures the purity and objectivity of the subsequent analysis of SAV and PT mode choice, allowing it to focus on travel demands where a genuine competitive relationship exists. This provides a clear and reliable baseline for evaluating their competition structure.

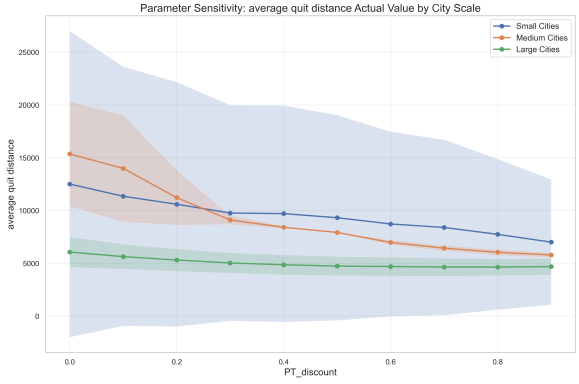
As illustrated in Figure 5.26c, the average distance of trips shifting from SAV to PT exhibits markedly different price elasticities across various city scales. In large cities, this distance decreases significantly as $\delta_{PT}$ increase, indicating that pricing policies are systematically altering the travel characteristics of attracted users. On the other hand, the quit distance in medium and small cities remains almost unaffected by the level of discount, showing a high degree of stability. To explain the underlying mechanisms of this phenomenon, it is necessary to further refine the classification of competitive environments across different cities.



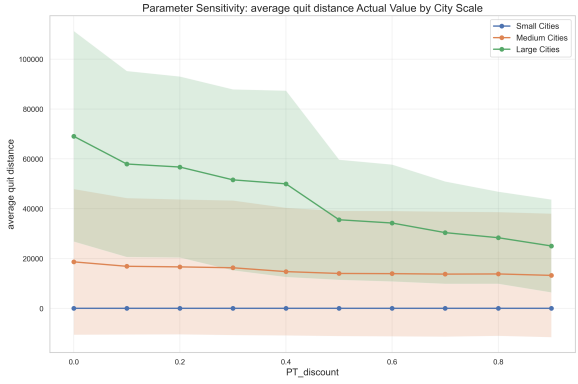

(a) PT mode share (Effective Competition Set)




(b) PT mode share (Network-Restricted Set)




(c) Avg. quit trip distance (Effective Competition Set)




(d) Avg. quit trip distance (Network-Restricted Set)

Figure 5.27: Impact of $\delta_{PT}$ across Different City Groups

The comparative analysis between the *Effective Competition Set* and the *Network-Restricted Set* (Figure 5.27) clearly reveals the decisive influence of network availability on price competition.

Observing the PT mode share (Figures 5.27a and 5.27b), the two groups of cities exhibit distinct responses. In the *Network-Restricted Set*, the PT mode share shows almost no response to price discounts. Particularly in small cities, the share remains consistently close to zero, which strongly supports the conclusion that the physical network is the fundamental bottleneck limiting PT's competitiveness. In contrast, within the *Effective Competition Set*, the PT mode share grows steadily with increasing discount levels, indicating that when network availability is adequate, price incentives are an effective means of enhancing its market competitiveness.

Figures 5.27c and 5.27d provides a fundamental explanation for the previously observed macro-level trends. A central finding is that the average quit distance in the *Network-Restricted Set* is substantially higher in absolute values than in the *Effective Competition Set*. This difference directly reveals the core competitive range between SAV and PT under the two market structures. In the *Network-Restricted Set*, the PT network is primarily oriented toward serving long-distance trunk corridors, resulting in effective competition with SAV only for long-distance trips. Consequently, SAV users who are attracted by discounts in these cities are essentially undertaking long-distance travel, which leads to the extremely high average quit distance observed in this group. In contrast, in the *Effective Competition Set*, the high spatial accessibility of the PT network enables comprehensive competition with SAV across all distance ranges. This not only explains the lower absolute value of the average quit distance in this group but also reveals more nuanced internal mechanisms. Within this group, the quit distance decreases with increasing discounts across all city scales. This strongly demonstrates that price discounts systematically attract users with shorter trips that previously found in large cities is a general feature of the *Effective Competition* environment. However, it is noteworthy that within the *Effective Competition Set*, the reduction in quit distance is more gradual in large cities compared to medium and small cities. The underlying reason is that the well-developed PT networks in these large cities, even before discounts are applied, have already established a strong competitive advantage and effectively captured the market for short- and medium-distance trips due to their high accessibility. Therefore, increasing discounts in these cities serves to enhance competitiveness more evenly across all distance ranges, rather than focusing on short-distance markets as in medium and small cities. As a result, the newly attracted users are more widely distributed in terms of trip distance, rather than being concentrated in short-distance segments, which naturally leads to a more moderate decline in the average quit distance.

To assess the relative competitiveness of PT across different distance segments, this study draws upon the concept of interaction effects used in the dual-parameter sensitivity analysis(Section 5.3.2) to construct a Public Transport Competitiveness Index. This index, $CI_i$, is defined for a specific distance bin $i$ as the ratio of two proportions:

$$CI_i = \frac{P_i}{D_i} = \frac{n_{i,PT}/N_{PT}}{n_i/N_{OD}} \tag{5.1}$$

where $n_i$ and $n_{i,PT}$ are the number of original trips and PT-switching trips in bin $i$ respectively, and $N_{OD}$ and $N_{PT}$ are their respective totals. It is crucial to note that the set of original trips, $N_{OD}$, is pre-filtered to exclude unreasonable trips by walking. This step ensures the analysis focuses exclusively on the market segment where motorized transport modes like SAV and PT are in direct competition.

The index intuitively measures the propensity of trips within a specific distance bin to switch to PT, relative to the overall average. A value of $CI_i > 1$ indicates that trips in this bin are more inclined than average to choose PT, signifying a relative advantage for PT in this market segment. Conversely, $CI_i < 1$ indicates a relative disadvantage, while $CI_i = 1$ represents a baseline level of competitiveness, where the propensity to switch is equal to the overall average.

Furthermore, the significance of this index is rooted in its direct relationship with the bin-specific switch rate, $R_i$, which is the proportion of trips within a bin that actually switch to PT ($R_i = n_{i,PT}/n_i$). By rearranging the terms from Equation 5.1, this relationship becomes explicit:

$$CI_i = \left(\frac{n_{i,PT}}{n_i}\right) \cdot \left(\frac{N_{OD}}{N_{PT}}\right) = R_i \cdot \frac{1}{R_{avg}} = \frac{R_i}{R_{avg}} \tag{5.2}$$

where $R_{avg} = N_{PT}/N_{OD}$ is the global average switch rate. This equation reveals that $CI_i$ is the ratio of the true switch rate in a specific distance bin, $R_i$, to the global average switch rate, $R_{avg}$. The baseline of $y = 1$ in the subsequent figures therefore represents the point where the switch rate equals the average, providing a clear benchmark for analysis.

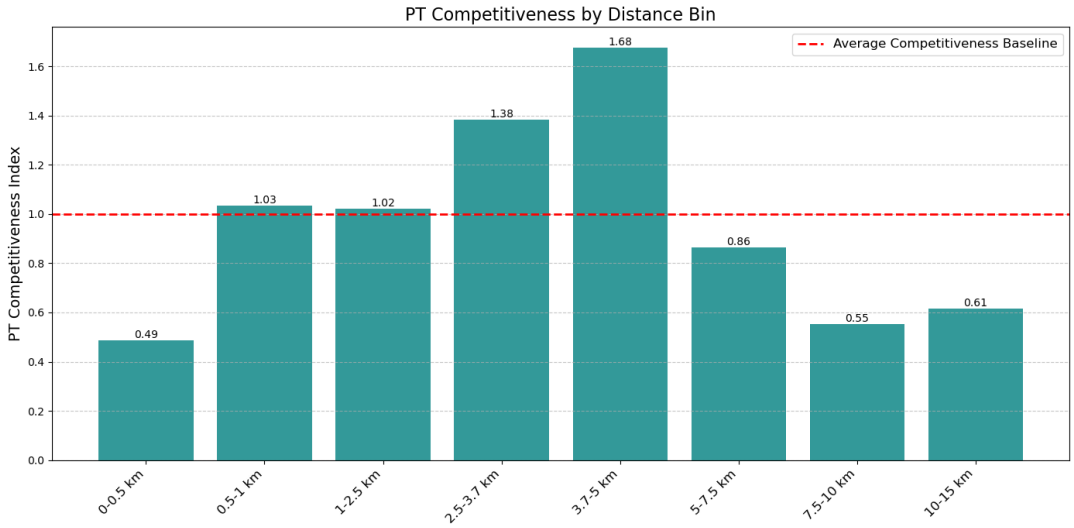
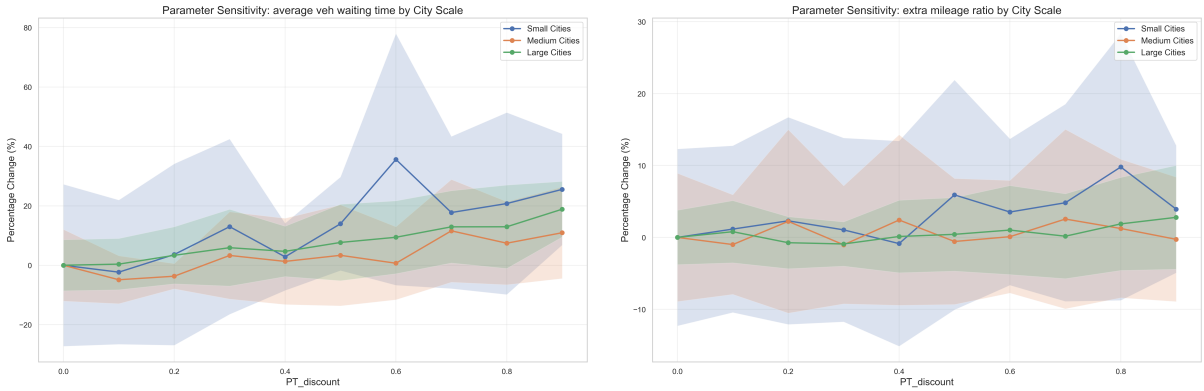



Figure 5.28: PT Competitiveness Index by Distance for Rotterdam

Figure 5.28 visualizes the PT Competitiveness Index for Rotterdam. The results reveal a complex, non-monotonic relationship between PT competitiveness and travel distance. For trips under 0.5 km, the index is low at 0.49. It then rises to approximately 1.0 for trips in the 0.5-2.5 km range, indicating a state of competitive balance between the two modes. PT's competitive advantage is strongest in the medium-distance range, building from the 2.5-3.7 km bin and peaking at 1.68 in the 3.7-5 km bin. In this segment, PT shows a decisive advantage over SAVs. However, for all distances beyond 5 km, PT's competitiveness sharply decreases, with the index falling well below 1.0. As trip duration increases, in-vehicle travel time becomes the dominant factor. The higher average travel speed defined for SAVs in the model leads to a significant cumulative time saving on longer routes. This inherent speed advantage ultimately outweighs PT's cost benefits, making SAVs the more competitive option for these extensive journeys. This non-monotonic competition structure is foundational to the trends observed in the *Effective Competition Set*, explaining both how increasing discounts lower the average quit distance by attracting shorter trips, and why this effect is more moderate in large cities where PT already dominates its most competitive segments.




(a) Average Vehicle Waiting Time

(b) Extra Mileage Ratio

Figure 5.29: Impact of PT Competition on SAV Operational Efficiency(Effective Competition Set)

The analysis extends to the impact of external competition on the SAV system's operational efficiency, focusing on the *Effective Competition Set* where these effects are most clear. As $\delta_{PT}$ increase, both the average SAV vehicle waiting time and its extra mileage ratio rise slightly (Figure 5.29a and 5.29b). This outcome stems directly from the reduced demand density, as a lower density of trip requests decreases the matching frequency for available vehicles, which restricts their service efficiency.

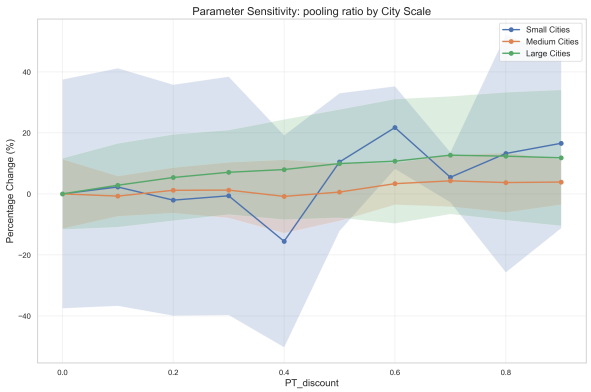



Figure 5.30: Impact of PT Competition on Pooling Ratio(Effective Competition Set)

Additionally, it is observed that as $\delta_{PT}$ increases, the pooling ratio of the SAV system rises. This is a structural effect of the nested Logit model's application in this study. For trips where both a shared and a solo ride are offered, they form an SAV nest. Given the high similarity ($\lambda = 0.3$), the nest's overall utility is enhanced by the more attractive shared ride option, making it robustly competitive against PT. In contrast, when only a solo ride is available, it competes directly with PT and is more likely to be diverted as $\delta_{PT}$ increases. This filtering of the demand leaves a higher concentration of trips suitable for pooling, which in turn increases the observed pooling ratio. In short, the pooling ratio rises because PT competition optimizes the internal structure of SAV demand.

### 5.5.1 Joint Effects of Congestion and Public Transport Discount

This section employs a two-parameter sensitivity analysis to systematically examine the joint effects of traffic congestion, measured by average SAV speed ($v_{avg}$), and the public transport pricing strategy ($\delta_{PT}$). This analysis seeks to understand the independent impact of each variable on competition structure and to determine the extent of any interaction between them.

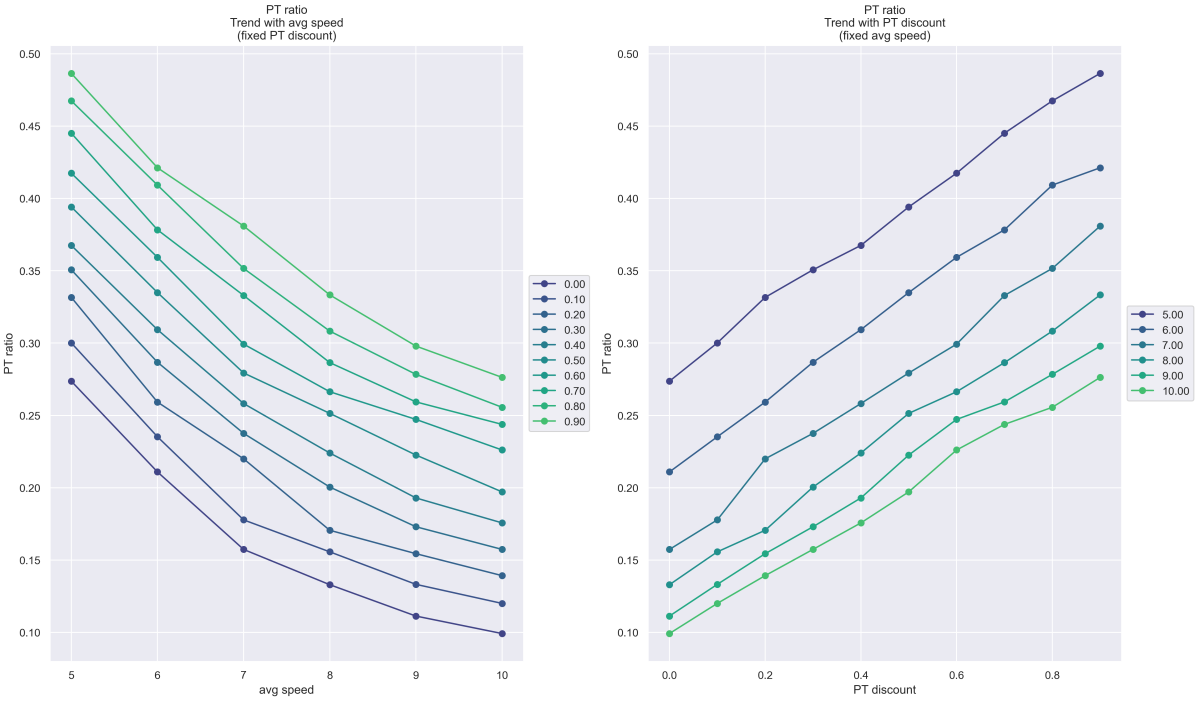



Figure 5.31: Interaction Effect of Average SAV Speed and $\delta_{PT}$ on PT Mode Share

As observed in the case of Rotterdam (Figure 5.31), the PT mode share systematically decreases as the average SAV speed increases, a trend that is consistent across all $\delta_{PT}$ levels. Notably, the slope of

the curves becomes less steep as speed increases, indicating that the PT mode share is more sensitive to changes in speed in low-speed (high-congestion) scenarios. This indicates that the negative impact of traffic congestion on the SAV system is amplified when PT has a price advantage.

A core feature of the SAV service is its door-to-door model, but its utility for users is highly dependent on travel time. When urban congestion leads to a decrease in average SAV speed, the total travel time for SAVs increases significantly, raising time costs and thus reducing their attractiveness. At the same time, the PT system, being unaffected by road traffic, maintains a relatively stable travel time. Therefore, as congestion worsens, users are more inclined to choose PT, and the increase in PT mode share is magnified. Conversely, in high-speed (low-congestion) scenarios, SAVs can fully leverage their time advantage, and the sensitivity of the PT mode share to changes in speed is significantly reduced.

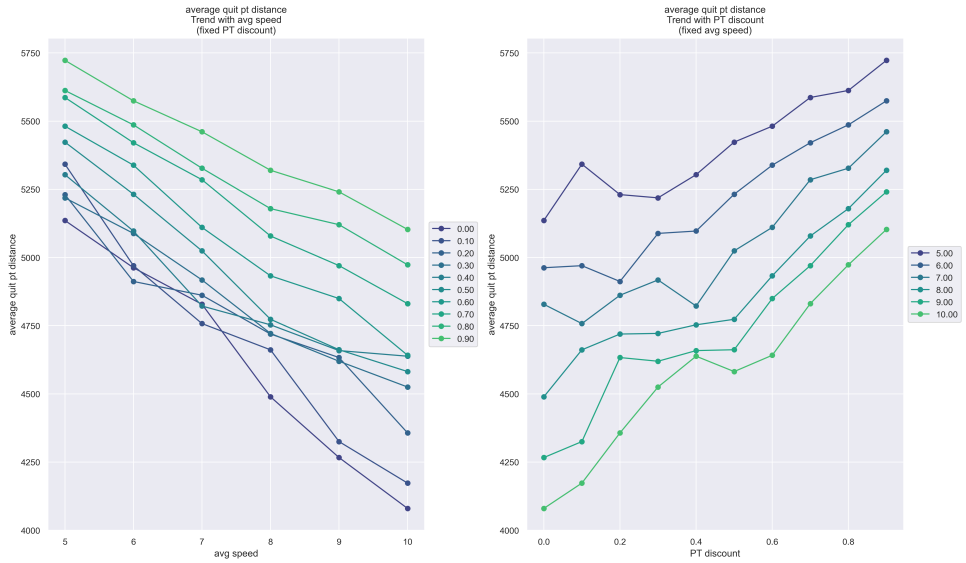



Figure 5.32: Average Distance of the PT Leg

The left panel of Figure 5.32 shows that as SAV operating speed increases, the average PT segment distance among switchers exhibits a moderate downward trend. This does not indicate the changing of the fundamental advantage of SAV in long-distance travel. Since long-distance trips are inherently the core advantage area for SAV, improvements in speed further reinforce this advantage, as the time savings accumulate proportionally with trip distance. This enhancement increases the relative utility of SAV for long-distance travel, enabling it to more effectively retain users who are particularly sensitive to travel time. As a result, the group of users switching to PT becomes more concentrated in medium and short-distance trips, leading to the observed moderate decline in average PT segment distance.

In contrast, as shown in the right panel, increasing $\delta_{PT}$ systematically raises the average PT segment distance among those who switch. This trend reflects the reshaping of the competition structure by discounts: by substantially reducing travel costs, discounts effectively strengthen PT's competitiveness in long-distance travel segments that were previously dominated by SAV's speed advantage, thereby attracting more long-distance users to switch and increasing the average PT segment distance within this group. It is important to note that this analysis focuses on the PT segment distance within switching trips, which is defined differently from the previously discussed origin request trip distance. A seemingly paradoxical phenomenon is that while higher $\delta_{PT}$ reduce the average total trip distance of switchers by attracting more short-distance users, they simultaneously increase the average PT segment distance by enhancing PT's competitiveness in long-distance travel. These two trends, though apparently contradictory, in fact reflect the model's differentiated impacts on various distance-based market segments under the complex trade-off between cost and time.

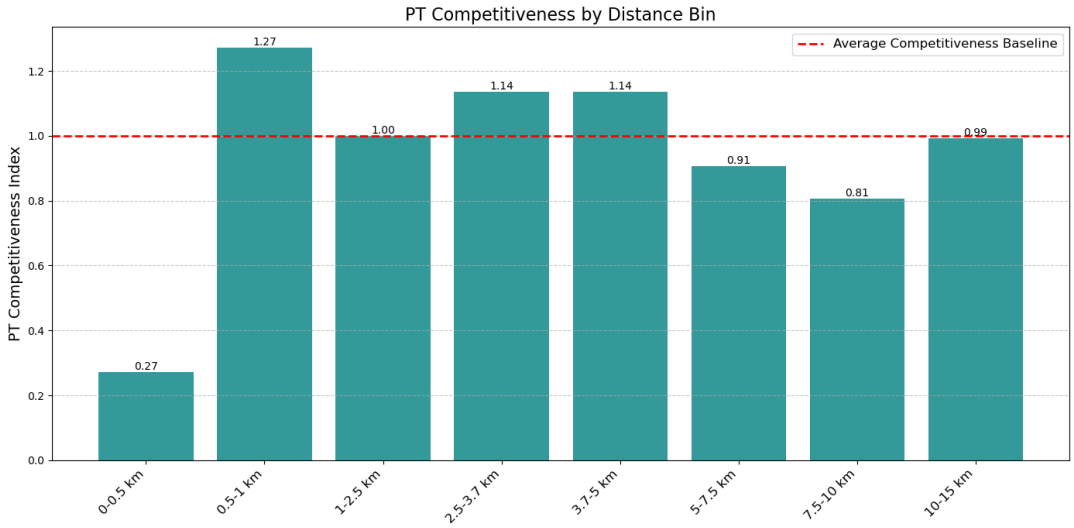
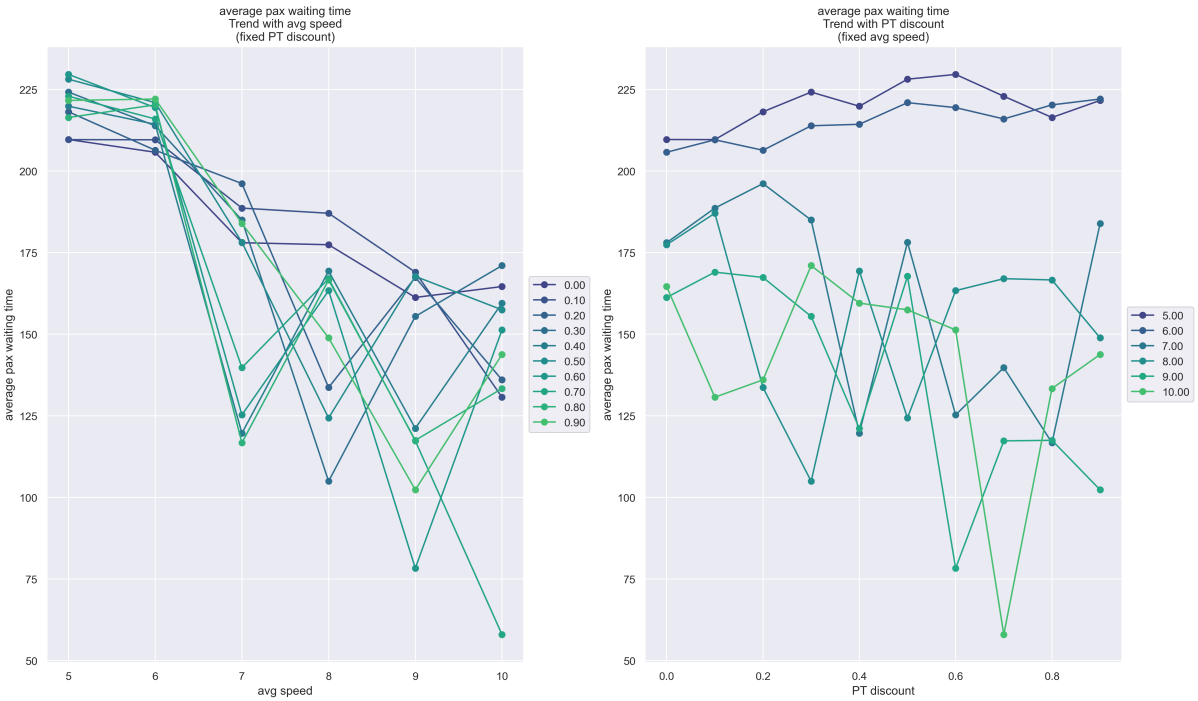


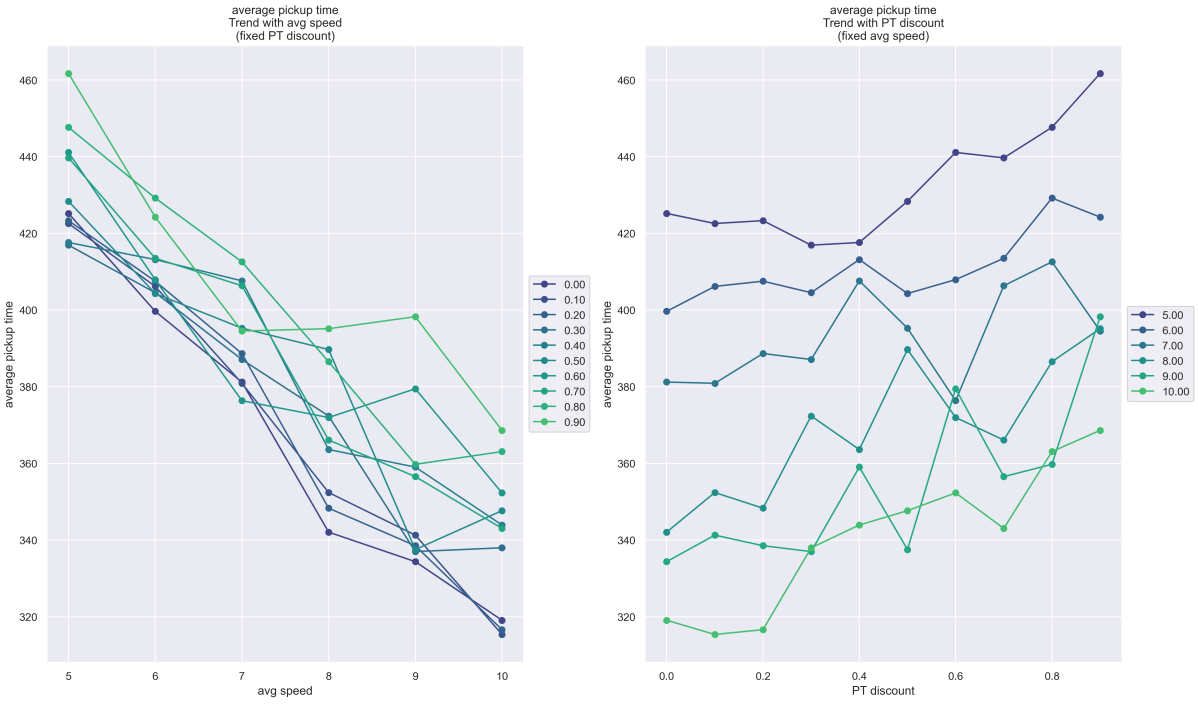

Figure 5.33: Competitiveness Index with a discount of 0.5

Taking a $\delta_{PT}$ rate of 0.5 as an example, Figure 5.33 illustrates the significant adjustment in the competitive structure between PT and SAV within specific market segments. However, compared to the baseline scenario, this pricing strategy does not fundamentally alter the overall competition structure; rather, it reshapes competitiveness within certain segments.

The application of $CI$ clarifies the effective scope and structural impact of PT's pricing strategy. For instance, in Rotterdam, even with a substantial 50% discount, the $CI$ for PT in the very short-distance market (<0.5 km) remains well below 1.0. This indicates that price incentives alone are insufficient to overcome PT's inherent disadvantages in last-mile convenience for the shortest trips. In contrast, the competition structure in the medium- and long-distance markets is significantly enhanced. The analysis reveals a key structural shift: the relative decline in the $CI$ for PT in the 2.5-5 km range does not signify a decrease in its attractiveness for these trips. Instead, it reflects the diminishing marginal utility of discounts in this segment, where PT already holds a strong advantage. Simultaneously, the discounts substantially boost PT's competitiveness in the long-distance market (>5 km), attracting a new cohort of long-distance travelers. This influx structurally alters the composition of PT switchers, causing the rise in the average PT segment distance to be primarily driven by these newly acquired long-distance trips.




(a) Average Passenger Waiting Time




(b) Average Pickup Time

Figure 5.34: Trends of Various SAV Indicators

In addition, this section examines how changes in external conditions affect the internal operational efficiency of the SAV system, as illustrated in Figure 5.34.

From the passenger perspective, average waiting time (Figure 5.34a) exhibits a highly nonlinear sensitivity to traffic congestion, with its response pattern attributable to two distinct mechanisms. In low-speed conditions (for example, when the average speed falls below 7 m/s), system operations are directly constrained by physical limitations. In this regime, vehicle travel speed becomes the critical bottleneck, causing pickup times to increase sharply as speed decreases, which in turn dominates total passenger waiting time. However, once congestion eases and speed surpasses a certain threshold, waiting

time becomes primarily determined by the efficiency of supply-demand matching. In this regime, passenger waiting time is primarily determined by dispatch time, which depends on the balance between instantaneous demand and available fleet size. This interval depends on system saturation, that is, the balance between instantaneous demand and available fleet size. Meanwhile, average pickup time (Figure 5.34b) displays a more linear trend across the entire speed range, as it more directly reflects the dominant role of physical travel efficiency in the SAV system.

As shown in the right panel, higher $\delta_{PT}$ lead to a systematic increase in average pickup time. This effect is primarily driven by reduced demand density: as some SAV users shift to PT, the remaining passengers are more widely dispersed, so vehicles must travel longer distances on average to reach each passenger. This results in lower fleet operational efficiency. However, this decline in efficiency does not necessarily imply a reduction in passenger service quality.

It is noteworthy that the interaction effect between traffic congestion and $\delta_{PT}$ is relatively weak. As shown in the figure, the performance curves under different $\delta_{PT}$ levels remain nearly parallel, indicating that changes in $\delta_{PT}$ do not fundamentally alter the direction or magnitude of the impact of speed on system performance. This suggests that the physical constraints imposed by congestion and the demand management induced by PT pricing operate as largely independent mechanisms affecting SAV system performance. Physical constraints, namely vehicle speed, determine the basic time required to complete tasks, while demand levels influence service efficiency by adjusting demand density.

# Chapter 6

# Discussion

This chapter provides a comprehensive synthesis of the study, beginning with a summary of the research scope and methodological contributions, followed by the main findings and their practical implications. It also offers actionable recommendations for operators, critically assesses the inherent limitations of the modeling approach, and concludes with suggestions for future research directions inspired by both the identified limitations and insights gained throughout the study.

## 6.1 Limitations

Although the framework developed in this study strives for comprehensiveness, it is still built upon a series of simplifying assumptions. Crucially, it must be noted that the underlying analysis identifies statistical correlations, not causal relationships; any interpretations of causality are based on logical reasoning within the model's framework. These limitations, detailed below, constitute the study's boundaries and clearly set the scope for the applicability of its conclusions.

### 6.1.1 Limitations of the Framework and Research Scope

Firstly, the most fundamental characteristic of the model is its static and offline nature, which is reflected in two aspects. During the generation of feasible pooling options, the core ExMAS algorithm operates offline: it assumes that all travel demands within a given time window are known in advance, allowing for a global search for optimal pooling combinations. This approach is well suited for strategic-level analysis of pooling potential. At the vehicle assignment stage, the simulation framework adopts a discrete-event assignment logic. The system processes trips sequentially according to their scheduled departure times, assigning the most suitable available vehicle to each ride. This process is myopic and does not optimize all vehicles and trips globally across time dimension.

The combination of offline matching and sequential assignment is a key feature of this framework. While it improves computational speed, it also means that the results are not globally optimal for vehicle-ride matching. Therefore, the ILP solver in the final step only finds an optimal solution for a reduced and pre-filtered set of vehicle-ride pairings generated by the sequential, non-optimal assignment stage. In addition, when evaluating competition with PT, although different congestion scenarios are introduced, the travel time for a given trip and departure time is deterministic and does not capture fluctuations caused by random events such as traffic accidents.

Secondly, one of the core limitations of this study arises from its simulation framework based on single-day demand. This inherently means that the results do not capture temporal heterogeneity, such as differences between weekdays and weekends or seasonal variations. For passengers, preferences such as willingness to share are assumed to be constant within a single day, so the model cannot reflect the evolution of preferences based on periodic travel experiences. For operators, the study lacks insights into scale effects that emerge over long-term operations, such as reductions in daily costs. As a result, the conclusions of this study are limited in their ability to explain long-term market evolution.

Another core simplification lies in the definition of the objective function. This study adopts a cost minimization paradigm at the methodological level. However, this approach does not cover two dimensions that are essential for a comprehensive evaluation of SAV systems. The first is the operator's profit model. In reality, operator decisions involve balancing pricing, service levels, cost control, and market demand, with the ultimate goal typically being profit maximization. The framework in this

study does not include revenue, so it cannot simulate the trade-off between costs and revenues. As a result, the findings mainly reflect the boundaries of operational efficiency under given demand, rather than the economic viability of a complete business model. The second dimension is social benefits and externalities. The value of ride-pooling services to urban transportation systems largely lies in their potential positive external effects, such as reducing total vehicle miles traveled, thereby alleviating congestion, lowering carbon emissions, and reducing air pollution. The current objective function does not monetize these externalities or include them in the optimization process.

In the analysis of the competitive relationship between SAVs and public transport, a notable limitation arises from the definition of the PT system boundary. In the current model, due to the availability of network data, the PT network is limited to rail systems only and does not include bus services as part of PT. This design constrains the scope of the conclusions. The study finds that in small and medium-sized cities with underdeveloped rail networks, PT does not present significant competition to the SAV system. To some extent, the results reflect the competition between SAVs and rail systems. However, considering that bus systems are the main component of public transport in many small and medium-sized cities, providing the primary network coverage and service, the current model cannot fully assess the competition between SAVs and the entire PT system.

### 6.1.2 Simplifications in Model Assumptions

The study adopts a single, monopolistic operator perspective and does not consider the involvement of private vehicles or operators. This overlooks the competition structure and equilibrium present in real markets. In reality, multiple SAV platforms may compete, leading to more complex passenger choice behaviors. The current model cannot capture such competitive market equilibria, so its conclusions are more applicable to a regulated market environment. In addition, the SAV fleet is treated as a homogeneous and perfectly controllable resource, ignoring real-world requirements such as vehicle charging or maintenance.

There are two levels of limitations regarding the treatment of walking in this study. Walking is not modeled as a fully equivalent travel mode alongside SAV and PT, but is instead filtered out for extremely short trips based on a utility comparison. In this mechanism, the utility of walking is mainly quantified by travel time, while the utility of SAV is a composite function of time, cost, and service level. As a result, this filter—originally intended to exclude only unreasonable SAV trips—may be restrictive and remove short trips that passengers might realistically choose SAV for due to comfort, safety, or convenience, potentially leading to an underestimation of the true demand and potential service value of SAVs in the short-distance market.

Similarly, the public transport system, as the main competitor to SAVs, is also highly simplified in the model. When calculating PT travel times, the model primarily relies on a macro-level estimate based on network distance and average speed. Specifically, it does not account for actual timetables or service frequencies, which means it cannot accurately simulate the real, time-varying waiting and transfer times at stations. Furthermore, the conclusions are based on the specific price structures assumed for SAV and PT, their universal applicability warrants further investigation. In addition, a uniform operating speed is used for all rail-based PT modes, which does not capture the heterogeneity between different modes or the availability and connectivity of different rail stations. While these simplifications ensure computational efficiency, they may lead to an underestimation of the total cost of PT travel and may not fully reflect the attractiveness or disadvantages of PT in mode choice competition with SAVs, potentially affecting the final mode split results.

Moreover, a core simplification of the simulation framework is its use of a fixed average speed for SAV operations, an approach that does not capture congestion effects arising from dynamic traffic flow. The rationale for this simplification lies in the architecture of the ExMAS algorithm, which is designed for efficient large-scale screening of trip pairs and relies heavily on a pre-generated, static travel time skim matrix. The core computational efficiency of the algorithm is derived from its vectorized calculation mechanism. Specifically, when identifying shareable trip pairs, the algorithm does not process each pair individually. Instead, all potential combinations are organized into a large data table, and fixed travel times from the skim matrix are attached in a single, high-performance batch operation. Introducing dynamic speeds would require each travel time calculation to be determined individually based on the departure time of each trip. This would undermine the vectorized foundation of the algorithm, forcing a shift from efficient batch processing to an iterative mode where each trip pair is computed separately. For large-level simulations, such a shift would result in orders-of-magnitude increases in computational cost, rendering the approach infeasible. Therefore, the adoption of a static speed model represents a trade-

off in ExMAS between computational feasibility at scale and model reality, and reflects a necessary compromise in realism for this study.

Furthermore, the selection and definition of the urban characteristic indicators themselves represent a form of simplification. For instance, the road network topology indicators are derived from specific graph theory-based abstractions of the physical road network. Different methods of network simplification or alternative topological metrics could potentially yield different correlation results, highlighting the model's dependence on these specific representational choices.

### 6.1.3 Endogenous Limitations of Dispatch Algorithms and Operational Strategies

The system does not employ explicit vehicle relocation mechanisms when assigning vehicles, but instead makes assignments based on vehicle availability and the maximum passenger waiting time threshold to realize vehicle relocation. It is important to note that if a vehicle is selected for a trip, its idle time is included in the waiting time for that trip. Because the total cost function includes a penalty for waiting time, the system is inherently incentivized to allocate vehicles that can provide continuous service, which inherently results in a distribution where some vehicles operate intensively while others remain idle. This results in an uneven utilization of vehicle resources. For vehicles that remain idle and are not assigned to any trips, the lack of explicit relocation mechanisms means their likelihood of being assigned decreases over time, until during peak periods, when vehicle shortages force their assignment to trips, sometimes resulting in high waiting times and idle rates under certain conditions.

In addition, during the vehicle assignment process, the mechanism of traversing trips by their request times can result in inefficient use of fleet resources. Specifically, since a single passenger may correspond to multiple ride options, each of these rides is initially assigned a vehicle. Although the final optimization ensures that each passenger is ultimately served by only one vehicle, this mechanism—where a passenger temporarily occupies multiple vehicles—may theoretically lead to a slight underestimation of overall fleet operating efficiency.

### 6.1.4 Dependence on Data Foundations and Parameter Calibration

The study uses a 1% sample of the total urban travel demand as input, which ensures computational feasibility but may not fully capture the nonlinear scale effects that only emerge at real-world demand densities. In addition, all demand data in this study are jointly calibrated using available datasets from the Netherlands and the OSM road network. This means that the quantitative findings are context-dependent and require careful recalibration before being generalized to other countries. In the grouped analysis of small and large cities, the sample size is relatively small, which may make the results sensitive to outliers and thus limit the generalizability and robustness of the conclusions. Furthermore, in the cross-city comparison, a uniform average vehicle speed (8 m/s) is set for all cities. This value is based on real data from large cities and is used to strictly control variables, allowing for a clear examination of the impact of urban structure itself on system performance. However, this also means the model may systematically underestimate SAV operating efficiency in small cities, where traffic conditions are typically better. Therefore, the conclusions of the cross-city analysis should be interpreted as structural differences under equivalent congestion levels, rather than as a direct reproduction of actual efficiency in each city.

There are also specific limitations at the parameter calibration level. First, the initial fleet size is set to the theoretical minimum required to meet all demand without triggering the generation of new vehicles. This means that all analyses are conducted under a background of maximized vehicle resource efficiency and zero redundancy. This is an idealized scenario and differs from real-world operations, where buffer capacity must be maintained to accommodate demand fluctuations and ensure service reliability. Second, the calibration of $c_w$ is primarily aimed at making the total vehicle waiting cost and total driving cost comparable in magnitude under the baseline scenario. This is based on the assumption that operators assign equal weight to both, rather than on a precise quantification of the real-world opportunity cost of vehicle waiting. In addition, the value of $\alpha$ is determined by observing the average trend across all studied cities, with the aim of conducting a macro-level analysis. However, given the significant heterogeneity between cities, this value may not represent the optimal strategic balance point for some cities.

Moreover, the resulting values of interaction effects are relative to the limited range; changing the parameter space would alter the calculated main and interaction effects, meaning the findings on synergistic or antagonistic effects are context-dependent and confined to the tested range.

Finally, a limitation arises from the interaction between VoT and $\omega$ during modeling and calibration. In theory, $VoT$ measures the general economic cost of travel time, while $\omega$ is intended to quantify the additional psychological cost weight associated with choosing shared rides. However, in this study, a uniform $VoT$ value is assigned to all passengers—regardless of commuting, non-commuting, in-vehicle, or waiting time—across all cities, without reflecting heterogeneity among passengers or between cities. As a result, the absolute value of $\omega$ obtained in this study cannot be simply interpreted as a perfect mapping of its theoretical definition. Instead, it represents a combined effect that includes both sharing preferences and unmodeled heterogeneity in $VoT$.

These model-based calibrations together define the baseline scenario of this study. As a result, the findings primarily provide strategic guidance for operators rather than precise numerical predictions, and practical applications should be adjusted and optimized according to the unique characteristics and realistic operational data of each city.

## 6.2 Practical Recommendations

This section synthesizes the main findings of the study and provides corresponding practical recommendations for operators. This discussion aims to revisit and synthesize these recommendations around five key themes: fleet sizing, operational efficiency, ride-pooling efficiency, market expansion, and competition with PT(public transport), all suggestions are based on the specific assumptions and boundaries defined in the previous section.

### 6.2.1 Fleet Sizing and Resource Allocation

At the resource allocation and fleet sizing level, the study reveals a moderating effect of city scale, indicating that a uniform resource allocation model may not yield optimal results across different urban contexts. The most important finding is that urban population size serves as the fundamental determinant for predicting the required total fleet size. In particular, in large cities, there is an almost perfectly linear positive correlation between required fleet size and population, making population the primary and decisive macro-level indicator for operators when planning market entry and capacity.

Furthermore, the results indicate that in order to achieve the same service level, small cities require a higher initial vehicle provision relative to demand compared to large cities. This is primarily because demand in small cities is more geographically dispersed, resulting in longer vehicle occupancy times per trip and consequently lower fleet turnover efficiency. Therefore, operators should avoid applying a uniform deployment ratio; when entering small city markets, it is necessary to increase the initial vehicle density in a targeted manner to ensure service reliability and availability.

Additionally, the spatial characteristics of travel demand, especially average commuting distance, are important factors in small cities. Where population size is similar, cities with longer average commuting distances require a larger fleet. For these cities, operators should allocate more vehicles to offset the decrease in operational efficiency.

### 6.2.2 Operational Efficiency Optimization

City scale is a key factor influencing both operational efficiency and service quality. Large cities, due to their high population and demand density, naturally exhibit greater operational efficiency, as reflected in lower extra mileage ratio and shorter average pickup times. In such environments, the operational focus should be on leveraging these density advantages. In contrast, small cities exhibit a less consistent service level, evidenced by a wider distribution of passenger waiting times. The primary challenge for operators in these contexts is therefore to enhance service reliability and consistency.

The topology of the road network is another set of key factors influencing system performance. For example, in small cities with high local clustering, as measured by the *average cluster coefficient*, vehicles have greater potential for reuse. This suggests that deploying capacity within tightly connected communities can facilitate faster and more continuous vehicle turnover. In medium-sized cities, road network density is closely associated with average pickup time. Therefore, An effective approach is to designate high-density areas as core operational zones, while implementing specialized dispatch strategies—such as dynamic pricing or reservation-based services—for low-density areas to address inherent efficiency limitations. In addition, a generally applicable finding relates to network complexity. In cities with more complex intersections, indicated by a higher *average degree*, there is greater potential to reduce

total vehicle mileage through efficient ride-pooling. Thus, promoting and optimizing ride-pooling route planning is a key strategy for reducing mileage-related operating costs.

Finally, policy choices regarding service parameters require clear strategic trade-offs. For example, increasing the $\Delta t^{p,\max}$ can lower vehicle idle costs but leads to higher empty mileage and longer passenger waiting times. The optimal setting depends on the main cost drivers: a shorter delay is preferable when fuel and maintenance costs dominate, while a longer delay may be suitable if parking fees or vehicle utilization are more critical. In all cases, decisions should be based on quantitative analysis to ensure that reduced idle costs sufficiently offset any increase in empty mileage and potential declines in passenger satisfaction.

### 6.2.3 Ride-Sharing Service Design and Pricing

Regarding the design and pricing of ride-sharing services, the results show that the effectiveness of ride-sharing discounts and willingness to share ($\omega$) varies with city size and user behavior. In large cities, passengers are more price-sensitive within a narrow range, requiring differentiated pricing: moderate discounts are effective for users with high willingness to share, while much higher discounts are needed for those less willing. In contrast, small and medium-sized cities require a generally higher discount to stimulate pooling. Across all contexts, a discount range of 10–30% is generally effective in incentivizing ride-sharing. When setting specific discounts, operators should strive to accurately identify and effectively utilize the sensitive discount intervals, while also being alert to the potential for negative utility traps. For example, for users with high $\omega$ (low tolerance for inconvenience), only when the discount reaches or exceeds a certain threshold in the sensitive interval (20% to 30%) can their willingness to share rides be significantly stimulated. On the other hand, for users with low $\omega$, a relatively high acceptance can already be achieved at lower discount levels. It is particularly important to note that excessively low discounts, for example below 10% to 15%, are sometimes not only ineffective, but may even produce negative joint effects for some users with middle $\omega$ values.

In addition to regional differences in pricing strategies, other urban characteristics also inform the optimization of ride-sharing services. In large cities, the study finds that both average commuting distance and the average degree of network nodes are positively correlated with ride-sharing rates. This suggests that cities with longer commuting distances are naturally well-suited for promoting ride-sharing. Designing targeted service strategies for users in these cities can more effectively leverage their inherent motivation to share rides.

Moreover, the focus of ride-sharing strategies should evolve with the stage of market development. In the early phase of market expansion, when demand density is low, the main operational challenge is to identify and match shareable trips within sparse demand. As the market matures and demand becomes denser, the core challenge shifts to effectively incentivizing potential users to accept ride-sharing. At this stage, passengers' personal preferences becomes the key determinant of ride-sharing success. Therefore, in mature markets, differentiated pricing and marketing systems based on user profiles become strategically essential.

### 6.2.4 Market Entry and Expansion

The central insight of this study lies in the pronounced scale effects observed in SAV (shared automated vehicles) services. Operators should fully recognize that the absolute values of key KPIs exhibit specific trends as demand scale grows. For instance, the average passenger waiting time shows limited correlation with demand volume and stabilizes within a consistent range, allowing operators to achieve a more precise balance between user experience and fleet operating costs based on the operational scope of specific cities.

Meanwhile, efficiency metrics such as average pickup time, average vehicle waiting time, and the extra mileage ratio generally demonstrate a scaling effect with increasing demand scale, showing improvement until reaching a relatively stable level. These trends of KPIs can support operators establishing reasonable expectations for service levels and fleet efficiency at different market development stages, thereby enabling the formulation of corresponding operational goals and resource allocation plans. In the early stages of market penetration, the complexity of ride-sharing rate performance requires significant attention from operators, as the interaction between passengers' willingness to share and demand scale is more significant, potentially leading to user behaviors that are not aligned with intuitive expectations. Thus, flexible and targeted operational adjustments are necessary during this phase. As the market scale gradually expands and the system moves beyond the initial complex interaction period, although passengers' own willingness to share remains critical and may even grow in its distinguishing role in ride-sharing be-

havior, necessitating continued targeted user management, the formulation and implementation of these strategies can be integrated into a more consistent operational structure. This is because the impact of further variations in demand scale on the effectiveness of these strategies becomes less significant. At the same time, operators should actively realize and utilize the scaling effects brought by demand growth to enhance fleet operational efficiency and optimize service costs. Finally, considering the saturation effect commonly observed in system performance improvement with demand scale growth, where marginal benefits diminish, operators should prioritize markets or regions still in lower demand density ranges where saturation effects have not yet been fully realized during resource allocation and market expansion to achieve higher returns on investment and performance gains. These analyses primarily offer strategic directional guidance for operational decisions, and their specific implementation should still be closely tailored and optimized based on the unique characteristics of each city and real-time operational data.

### 6.2.5 Competition Strategies with Public Transport

Competition strategies with public transport are multidimensional, with the core requirement being precise adjustment according to the local market environment. Fundamentally, the basis of competition is determined by the infrastructure coverage of the local PT network. In areas where the PT network is underdeveloped, SAVs can focus on optimizing internal operating efficiency. Conversely, in markets with a mature PT network, SAVs need to adopt direct competition strategies, which typically involve the use of price incentives.

A central finding for SAV operators is that PT pricing is not merely a competitive force that influences market share, but a significant mechanism that reshapes the structure of demand for their services. As the preceding analyses have shown, these pricing policies do not affect all SAV trips uniformly. Instead, they alter the competition structure across different distance segments.

Travel distance is a key factor shaping the competitive relationship between SAVs and PT. The analysis shows that SAVs maintain a structural advantage in the short-distance market, where door-to-door convenience and reduced access time are decisive. As trip distance increases, PT's competitiveness does not simply increase monotonically; instead, it is most pronounced in the medium to long-distance segments (such as 5–15 km), where fare structures and network efficiency allow PT to attract a larger share of demand. However, in very long-distance travel, the speed advantage of SAVs can offset PT's cost benefits, limiting PT's dominance. For SAV operators, it is crucial to recognize the growing competitive pressure from PT in the medium to long-distance market, especially as PT pricing policies or discounts are introduced. In these segments, strategies such as dynamic pricing, targeted service differentiation, or even collaboration with PT may be necessary to retain core user groups and respond effectively to shifting demand patterns.

Furthermore, competition from PT may lead to a slight decline in certain operational indicators due to reduced demand density, which should be regarded as an inherent outcome of market competition. The strategic focus can shift toward dynamically reallocating fleet resources to segments where SAVs have a comparative advantage, such as short-distance trips or areas with insufficient PT coverage, thereby maximizing efficiency under the new market equilibrium.

# Chapter 7

# Conclusions

This study has developed a comprehensive simulation framework to systematically investigate fleet management strategies for SAVs(shared automated vehicles) across 37 Dutch cities of varying sizes. By integrating an advanced ride-pooling generation algorithm with an innovative two-stage vehicle assignment optimization process, and explicitly modeling competition with public transport, the study provides in-depth insights into the complex interactions among urban characteristics, operational parameters, and system performance. The results address the core research question of how SAV fleet management should be adjusted in a multimodal urban transport context to balance operator and user costs.

This study contributes to the field of urban transportation by systematically addressing the heterogeneity of urban environments in SAV operations. Unlike most research that focuses on a single city, this work analyzes 37 Dutch cities, providing generalizable insights into how city size, road network topology, and demand patterns influence SAV system performance. A key contribution lies in explicitly simulating the competition between SAVs and public transport, offering a more realistic perspective for assessing the feasibility of SAVs within multimodal transportation systems. The developed integrated analytical framework bridges the gap between theoretical ride-pooling potential and practically executable, conflict-free vehicle scheduling, delivering a robust tool for strategic fleet management.

In terms of methodology, the main methodological contribution lies in the development of an innovative simulation framework to address sub-question 4, which enhances the ExMAS algorithm through a two-stage vehicle assignment process. In the first stage, a filtering mechanism based on discrete-event simulation is employed to efficiently generate a feasible decision space for vehicle-ride matching, substantially reducing computational complexity. The second stage applies integer linear programming to determine the globally optimal assignment scheme that minimizes the joint cost function of operators and users. This integrated approach links attractive ride-pooling solutions with executable, conflict-free vehicle schedules, thereby addressing key limitations of previous models. In addition, the framework incorporates a nested Logit model to simulate mode choice, with particular emphasis on the detailed construction of public transport travel chains. This approach enables a more accurate representation of multimodal travel behavior and supports quantitative analysis of market competition between SAVs and public transport.

## 7.1 Key Findings

Regarding sub-question 1, the study concludes that population size is the primary determinant of required fleet size, exhibiting a near-perfect linear correlation (correlation coefficient of 0.99) with the number of vehicles in large cities. However, for operational efficiency, more nuanced indicators are critical. In large cities, ride-pooling potential is driven by longer average commuting distances and higher network complexity, with both *average commuting distance* and *average degree* showing strong positive correlations (0.88 and 0.81, respectively) with the pooling ratio. In medium-sized cities, operational efficiency is more closely tied to network accessibility, where higher *road density* significantly reduces *average pickup times* (correlation of -0.54), thereby enhancing service responsiveness. Conversely, in small cities, where population scale is less of a differentiating factor, vehicle turnover efficiency is most influenced by the local compactness of the network—evidenced by a strong positive correlation (0.84) between the *average clustering coefficient* and the *vehicle reuse rate*—and by demand density, where *population density* is positively correlated (0.75) with the *average number of rides per vehicle*.

Regarding sub-question 2, the findings reveal that the effectiveness of pricing strategies is strongly modulated by city scale, exposing complex interaction effects between fare discounts and passenger resistance to sharing. A uniform discount strategy is ineffective because system sensitivity to these parameters varies significantly. Large cities, which possess high endogenous pooling potential, exhibit high sensitivity to pricing and resistance to sharing, with the analysis identifying specific parameter zones of strong synergistic or antagonistic interaction. This creates both opportunities and risks. For instance, for users with a high willingness to share, a discount of around 20% is highly effective as it falls within a zone of strong positive interaction. Conversely, for users with low sharing willingness, a 30% discount can be inefficient and unstable, falling into a zone of negative interaction despite high sensitivity; a higher discount of 50% or more is required to move into a stable, high-pooling regime. In contrast, small and medium-sized cities demonstrate lower sensitivity and weaker interaction effects, indicating that price-based incentives yield diminishing returns. In these contexts, achieving high pooling rates requires generally higher discounts (e.g., a 10-30% range proves only broadly effective) and likely needs to be supplemented with non-price strategies that enhance service quality.

Regarding sub-question 3, the research confirms the existence of significant scale effect. System efficiency, measured by metrics like *average pickup time* and *extra mileage ratio*, improves markedly with increasing demand but eventually reaches saturation. This reveals a critical shift in operational challenges: in nascent markets with sparse demand (e.g., a 1% demand scale), the bottleneck is the system's ability to find matches, resulting in a strong interaction between demand scale and passenger preferences. In mature markets with denser demand, matching opportunities are abundant, and the bottleneck shifts to incentivizing user choice, where passenger preferences become the dominant, independent factor constraining pooling adoption.

Regarding sub-question 5, the study concludes that the competitive relationship between SAVs and public transport (PT) is fundamentally governed by the structure and coverage of the PT network, rather than a simple speed advantage. This is evidenced by dividing cities into an 'Effective Competition Set' with dense PT networks and a 'Network-Restricted Set' with sparse networks. The analysis reveals that the average travel distance of trips switching to PT is significantly shorter in the effective set, demonstrating that well-developed PT networks compete across a broad spectrum of distances, whereas sparse networks can only compete on specific long-distance corridors. This leads to a non-monotonic competitive dynamic that is heavily dependent on travel distance. A developed *Public Transport Competitiveness Index* for a representative large city (from the effective set) quantifies this: SAVs hold a competitive advantage on short trips (<0.5 km), followed by a state of competitive balance (0.5-2.5 km). PT's advantage is dominant in the medium-distance range, peaking between 3.7-5 km. For trips over 5 km, SAVs regain dominance as it becomes less likely that the PT network's topology can provide a direct, efficient path for arbitrary origin-destination pairs, making the door-to-door flexibility of SAVs the decisive factor. Price-based competition from PT primarily enhances its competitiveness in the medium-to-long distance segments where its network coverage is effective. This selective filtering of demand has the structural effect of increasing the SAV system's internal pooling ratio, as solo SAV trips are more likely to be diverted to PT than trips for which an attractive shared SAV option exists.

## 7.2 Recommendations for Future Research

This study provides a foundational framework for understanding the operational patterns and challenges of SAV systems in different urban environments. Based on the findings and limitations, future research can be advanced along four dimensions: expanding the application boundaries of the model, deepening the complexity of operational strategies, enhancing the behavioral realism of the model, and conducting comprehensive assessments of socioeconomic impacts.

The first direction for future research is to extend the analysis from single-city settings to intercity and regional travel networks. Both this study and the vast majority of current SAV research fundamentally focus on a single city as the core unit of analysis. For example, Huang et al. (2022) proposed the coordinated optimization of dynamic ride-pooling and train timetables, achieving efficient mode integration based on the last-mile concept. However, the main contribution of such studies lies in optimizing SAV connections to transport hubs within city boundaries. A notable research gap remains regarding the role and value of SAVs in the complete travel chain from origin city to intercity transport to destination city. Therefore, future research should address the integration of SAV and PT at the regional level, rather than restricting analysis to interactions within a single city. In addition, it is important to develop models capable of evaluating scenarios in small cities with limited public transport networks, such

as assessing the extent to which SAV services can or should replace traditional fixed-route bus systems to more efficiently connect to regional transport networks. Achieving this goal also calls for significant improvements in the fidelity of PT models, including the incorporation of actual bus and rail timetables, service frequencies, and realistic transfer waiting times.

In terms of operational strategies and dispatch algorithms, future research should explore more complex models. Vehicle relocation is key to improving operational efficiency. Future work could integrate various relocation strategies from existing research, such as global optimization and demand forecasting (Fagnant and Kockelman, 2014), and test them in more complex competitive environments. In addition, it is worth exploring reservation-based operational paradigms, which are fundamentally different from the current real-time response mode and shift the optimization problem from myopic matching to a more global perspective. Furthermore, one of the core limitations of this study is the assumption of a single dominant operator in the market. A key direction for future research is to build hybrid market simulation frameworks that combine the centralized SAV optimization algorithm developed in this study with agent-based models of independent driver behavior (as demonstrated by the MaaSSim simulation library (Kucharski and Cats, 2022)), in order to investigate market equilibrium and optimal operational strategies when centralized fleets and competing ride-hailing platforms coexist.

In the area of pricing strategies, this study has proposed a simple discriminatory pricing approach in the analysis of passengers preferences. By evaluating the effects of different discount levels for passengers with varying willingness to share, the analysis provides insights into how targeted pricing strategies can be designed to maximize ride-pooling adoption. Future research can build on existing literature to consider a more comprehensive passenger heterogeneity (Rahman and Thill, 2025; van Engelen et al., 2024). Future work could further extend this approach by implementing and testing more refined discriminatory pricing schemes, such as personalized discounts based on a broader set of user attributes, or dynamic adjustment of discounts in response to real-time demand, supply, and user behavior. Incorporating more comprehensive user profiling and behavioral data would enable a more accurate and equitable assessment of discriminatory pricing in ride-pooling services. In addition, it is important to explore the fairness and ethical implications of discriminatory pricing in SAV systems.

Another promising direction is to establish effective topological and service level metrics for multimodal transport networks themselves. Future research should focus on building integrated multilayer network models and developing targeted topological indicators. For example, classical concepts such as betweenness centrality and closeness centrality may not be applicable in multimodal transport networks; thus, it is necessary to adjust or reconstruct these metrics, develop new indicators that capture the topology features of multimodal networks, and construct time-based comprehensive accessibility metrics. These efforts will enable a deeper understanding of the relationship between network structure and system efficiency or travel mode choices.

Moreover, at the operational environment level, a promising direction is to use mathematical models to capture the dynamic impact of demand characteristics on vehicle speed (for example, whether different trip distances may lead to varying congestion levels), differences in service levels across road types, and the feedback effects of SAV fleets on congestion. At the level of operational constraints, the model should incorporate more real-world complexities, such as SAV charging times and charging infrastructure layout, long-term seasonal demand fluctuations, as well as cost functions based on real operator data and more refined penalty models for waiting costs.

Finally, future research should move from using operational efficiency merely as a performance indicator to directly incorporating broader socioeconomic impacts into the optimization objectives. While the quantification of externalities such as air pollution and noise has been discussed in the literature (Schröder and Kaspi, 2024), a key direction is to incorporate these external costs directly into the optimization objectives of SAV system models. This would enable a more integrated analysis of how SAV operations can be optimized not only for operator profit but also for overall social welfare, thereby providing more robust support for sustainable urban mobility decisions.

# References


Niels Agatz, Alan Erera, Martin Savelsbergh, and Xing Wang. Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research*, 223(2):295–303, 2012. ISSN 0377-2217. doi:10.1016/j.ejor.2012.05.028. URL https://www.sciencedirect.com/science/article/pii/S0377221712003864.

Niels A.H. Agatz, Alan L. Erera, Martin W.P. Savelsbergh, and Xing Wang. Dynamic ride-sharing: A simulation study in metro atlanta. *Transportation Research Part B: Methodological*, 45(9):1450–1464, 2011. ISSN 0191-2615. doi:10.1016/j.trb.2011.05.017. URL https://www.sciencedirect.com/science/article/pii/S0191261511000671. Select Papers from the 19th ISTTT.

María J. Alonso-González, Niels van Oort, Oded Cats, Sascha Hoogendoorn-Lanser, and Serge Hoogendoorn. Value of time and reliability for urban pooled on-demand services. *Transportation Research Part C: Emerging Technologies*, 115:102621, 2020. ISSN 0968-090X. doi:10.1016/j.trc.2020.102621. URL https://www.sciencedirect.com/science/article/pii/S0968090X1931589X.

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. doi:10.1073/pnas.1611675114. URL https://www.pnas.org/doi/abs/10.1073/pnas.1611675114.

Theo A Arentze and Harry J.P Timmermans. A learning-based transportation oriented simulation system. *Transportation Research Part B: Methodological*, 38(7):613–633, 2004. ISSN 0191-2615. doi:10.1016/j.trb.2002.10.001. URL https://www.sciencedirect.com/science/article/pii/S0191261503000948.

Peyman Ashkrof, Gonçalo Homem de Almeida Correia, Oded Cats, and Bart van Arem. Understanding ride-sourcing drivers' behaviour and preferences: Insights from focus groups analysis. *Research in Transportation Business & Management*, 37:100516, 2020. ISSN 2210-5395. doi:10.1016/j.rtbm.2020.100516.

Milos Balac, Sebastian Hörl, and Kay W. Axhausen. Fleet sizing for pooled (automated) vehicle fleets. *Transportation Research Record: Journal of the Transportation Research Board*, 2674(9):168–176, 2020. doi:10.1177/0361198120927388.

Geoff Boeing. Modeling and analyzing urban networks and amenities with OSMnx. *Geographical Analysis*, 2025. doi:10.1111/gean.70009. Published online ahead of print.

F. T. Boesch and J. F. Gimpel. Covering points of a digraph with point-disjoint paths and its application to code optimization. *J. ACM*, 24(2):192–198, April 1977. ISSN 0004-5411. doi:10.1145/322003.322005.

Patrick M. Boesch, Francesco Ciari, and Kay W. Axhausen. Autonomous vehicle fleet sizes required to serve different levels of demand. *Transportation Research Record: Journal of the Transportation Research Board*, 2542(1):111–119, 2016. doi:10.3141/2542-13.

TomTom International BV. Tomtom traffic index: Ranking 2024. https://www.tomtom.com/traffic-index/ranking/?country=NL, 2024. Accessed: 2024-07-15.

Patrick M. Bösch, Felix Becker, Henrik Becker, and Kay W. Axhausen. Cost-based analysis of autonomous mobility services. *Transport Policy*, 64:76–91, 2018. ISSN 0967-070X. doi:10.1016/j.tranpol.2017.09.005. URL https://www.sciencedirect.com/science/article/pii/S0967070X17300811.

Oded Cats, Rafal Kucharski, Santosh Rao Danda, and Menno Yap. Beyond the dichotomy: How ride-hailing competes with and complements public transport. *PLOS ONE*, 17(1):1–17, 01 2022. doi:10.1371/journal.pone.0262496. URL https://doi.org/10.1371/journal.pone.0262496.

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, 2022. URL https://arxiv.org/abs/2207.14246.

Daniel J. Fagnant. Shared autonomous vehicles: Model formulation, sub-problem definitions, implementation details, and anticipated impacts. *2015 American Control Conference (ACC)*, pages 2593–2593, 2015. doi:10.1109/ACC.2015.7171124.

Daniel J. Fagnant. Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in austin, texas. *Transportation*, 45:143–158, 2018. doi:10.1007/S11116-016-9729-Z.

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. ISSN 0968-090X. doi:10.1016/j.trc.2013.12.001.

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. doi:10.1016/j.trc.2023.104398.

Andrés Fielbaum and Baiba Pudāne. Are shared automated vehicles good for public- or private-transport-oriented cities (or neither)? *Transportation Research Part D: Transport and Environment*, 136: 104373, 2024. ISSN 1361-9209. doi:10.1016/j.trd.2024.104373. URL https://www.sciencedirect.com/science/article/pii/S1361920924003304.

John Forrest, Ted Ralphs, Stefan Vigerske, Haroldo Gambini Santos, Lou Hafer, Bjarni Kristjansson, Miles Lubin, Matthew Saltzman, et al. coin-or/cbc: Release releases/2.10.12, May 2024. URL https://doi.org/10.5281/zenodo.13347261.

Masabumi Furuhata, Maged Dessouky, Fernando Ordóñez, Marc-Etienne Brunet, Xiaoqing Wang, and Sven Koenig. Ridesharing: The state-of-the-art and future directions. *Transportation Research Part B: Methodological*, 57:28–46, 2013. ISSN 0191-2615. doi:10.1016/j.trb.2013.08.012. URL https://www.sciencedirect.com/science/article/pii/S0191261513001483.

Marvin Greifenstein. Factors influencing the user behaviour of shared autonomous vehicles (savs): A systematic literature review. *Transportation Research Part F: Traffic Psychology and Behaviour*, 100: 323–345, 2024. ISSN 1369-8478. doi:10.1016/j.trf.2023.10.027.

Charles R. Harris, K. Jarrod Millman, Stéfan J van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585:357–362, 2020. doi:10.1038/s41586-020-2649-2.

Andreas Horni, Kai Nagel, and Kay W. Axhausen, editors. *The Multi-Agent Transport Simulation MATSim*. Ubiquity Press, London, 2016. doi:http://dx.doi.org/10.5334/baw. URL https://doi.org/10.5334/baw. License: CC-BY 4.0.

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. doi:https://doi.org/10.1016/j.compenvurbsys.2021.101730. URL https://www.sciencedirect.com/science/article/pii/S019897152100137X.

Michael F. Hyland and Hani S. Mahmassani. Taxonomy of shared autonomous vehicle fleet management problems to inform future transportation mobility. *Transportation Research Record*, 2653(1):26–34, 2017. doi:10.3141/2653-04.

Sebastian Hörl and Milos Balac. Introducing the eqasim pipeline: From raw data to agent-based transport simulation. *Procedia Computer Science*, 184:712–719, 01 2021. doi:10.1016/j.procs.2021.03.089.

Yutaka Ishibashi and Eizo Akiyama. Predicting the impact of shared autonomous vehicles on tokyo transportation using matsim. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 3258–3265, 2022. doi:10.1109/BigData55660.2022.10021068.

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. doi:10.1016/j.trc.2021.103236.

KiM Netherlands Institute. New values of travel time, reliability and comfort in the Netherlands, 2024. URL `https://doi.org/10.17026/SS/VSOOFW`.

Rafał Kucharski and Oded Cats. Exact matching of attractive shared rides (exmas) for system-wide strategic evaluations. *Transportation Research Part B: Methodological*, 139:285–310, 2020. ISSN 0191-2615. doi:10.1016/j.trb.2020.06.006.

Rafał Kucharski and Oded Cats. Simulating two-sided mobility platforms with maassim. *PLOS ONE*, 17(6):1–20, 06 2022. doi:10.1371/journal.pone.0269682. URL `https://doi.org/10.1371/journal.pone.0269682`.

Yusuke Kumakoshi, Hisatomo Hanabusa, and Takashi Oguchi. Impacts of shared autonomous vehicles: Tradeoff between parking demand reduction and congestion increase. *Transportation Research Interdisciplinary Perspectives*, 12:100482, 2021. ISSN 2590-1982. doi:10.1016/j.trip.2021.100482.

Ao Liu, Shaopeng Zhong, Daniel Sun, Yunhai Gong, Meihan Fan, and Yan Song. Joint optimal pricing strategy of shared autonomous vehicles and road congestion pricing: A regional accessibility perspective. *Cities*, 146:104742, 2024. ISSN 0264-2751. doi:10.1016/j.cities.2023.104742. URL `https://www.sciencedirect.com/science/article/pii/S0264275123005541`.

Carlos Llorca and Rolf Moeckel. Effects of scaling down the population for agent-based traffic simulations. *Procedia Computer Science*, 151:782–787, 2019. ISSN 1877-0509. doi:10.1016/j.procs.2019.04.106. URL `https://www.sciencedirect.com/science/article/pii/S1877050919305691`. 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.

Patricia C. Melo and Daniel J. Graham. Transport-induced agglomeration effects: Evidence for us metropolitan areas. *Regional Science Policy and Practice*, 10(1):37–48, 2018. ISSN 1757-7802. doi:10.1111/rsp3.12116. URL `https://www.sciencedirect.com/science/article/pii/S1757780223005814`.

Aitan M. Militão and Alejandro Tirachini. Optimal fleet size for a shared demand-responsive transport system with human-driven vs automated vehicles: A total cost minimization approach. *Transportation Research Part A: Policy and Practice*, 151:52–80, 2021. doi:10.1016/j.tra.2021.07.004.

Stuart Mitchell, Michael O'Sullivan, and Iain Dunning. PuLP: A Linear Programming Toolkit for Python. Technical report, Department of Engineering Science, The University of Auckland, September 2011. Available at: `https://optimization-online.org/?p=11731`.

Cristiano Martins Monteiro, Cláudia A. Soares Machado, Mariana De Oliveira Lage, Fernando Tobal Berssaneti, Clodoveu A. Davis, and José Alberto Quintanilha. Optimization of carsharing fleet size to maximize the number of clients served. *Computers, Environment and Urban Systems*, 87:101623, 2021. doi:10.1016/j.compenvurbsys.2021.101623.

Elaine M. Murtagh, Jacqueline L. Mair, Elroy Aguiar, Catrine Tudor-Locke, and Marie H. Murphy. Outdoor walking speeds of apparently healthy adults: A systematic review and meta-analysis. *Sports Medicine*, 51(1):125–141, 01 2021. ISSN 1179-2035. doi:10.1007/s40279-020-01351-3. URL `https://doi.org/10.1007/s40279-020-01351-3`.

Santhanakrishnan Narayanan, Emmanouil Chaniotakis, and Constantinos Antoniou. Shared autonomous vehicle services: A comprehensive review. *Transportation Research Part C: Emerging Technologies*, 111:255–293, 2020. ISSN 0968-090X. doi:10.1016/j.trc.2019.12.008.

OpenStreetMap. Openstreetmap [data set]. `https://www.openstreetmap.org`, 2024. Available as open data under the Open Data Commons Open Database License (ODbL).

ProRail. Prorail: Homepage, 2025. URL https://www.prorail.nl/. Accessed: 2025-5-28.

Boting Qu, Linran Mao, Zhenzhou Xu, Jun Feng, and Xin Wang. How many vehicles do we need? fleet sizing for shared autonomous vehicles with ridesharing. *IEEE Transactions on Intelligent Transportation Systems*, 23(9):14594–14607, 2022. doi:10.1109/TITS.2021.3130749.

Md. Mokhlesur Rahman and Jean-Claude Thill. What drives the use of pooled autonomous vehicles? some insights in california users' perspective. *Travel Behaviour and Society*, 39:100975, 2025. ISSN 2214-367X. doi:https://doi.org/10.1016/j.tbs.2024.100975. URL https://www.sciencedirect.com/science/article/pii/S2214367X24002382.

Jean-Paul Rodrigue. *The Geography of Transport Systems*. Routledge, London, 6 edition, 2024. ISBN 9781003343196. doi:10.4324/9781003343196. eBook, 402 pages.

Mauro Salazar, Federico Rossi, Maximilian Schiffer, Christopher H. Onder, and Marco Pavone. On the interaction between autonomous mobility-on-demand and public transportation systems. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 2262–2269, 2018. doi:10.1109/ITSC.2018.8569381.

Daniel Schröder and Mor Kaspi. Quantifying the external costs of autonomous on-demand ride pooling services. *Case Studies on Transport Policy*, 18:101302, 2024. ISSN 2213-624X. doi:10.1016/j.cstp.2024.101302. URL https://www.sciencedirect.com/science/article/pii/S2213624X24001573.

Toru Seo and Yasuo Asakura. Multi-objective linear optimization problem for strategic planning of shared autonomous vehicle operation and infrastructure design. *IEEE Transactions on Intelligent Transportation Systems*, 23(4):3816–3828, 2022. doi:10.1109/TITS.2021.3071512.

Jaime Soza-Parra, Rafał Kucharski, and Oded Cats. The shareability potential of ride-pooling under alternative spatial demand patterns. *Transportmetrica A Transport Science*, 20(2), 2024. ISSN 2324-9935. doi:10.1080/23249935.2022.2140022.

Sehyun Tak, Soomin Woo, Sungjin Park, and Sunghoon Kim. The city-wide impacts of the interactions between shared autonomous vehicle-based mobility services and the public transportation system. *Sustainability*, 13(12):6725, 2021. ISSN 2071-1050. doi:10.3390/su13126725.

Milan van Engelen, Vincent van den Berg, Eric Molin, and Bert van Wee. Understanding preferences for on-demand shared ride-hailing services: The role of travel time, pricing and socio-demographic characteristics. *Transportation*, 51(3):831–852, 2024. ISSN 1572-9435. doi:10.1007/s11116-023-10442-9. URL https://doi.org/10.1007/s11116-023-10442-9. Open Access.

Centraal Bureau voor de Statistiek (CBS) and Rijkswaterstaat (RWS). Onderzoek Verplaatsingen in Nederland 2014 - OViN 2014, 2015. URL https://doi.org/10.17026/dans-x95-5p7y.

Reza Vosooghi, Jakob Puchinger, Marija Jankovic, and Anthony Vouillon. Shared autonomous vehicle simulation and service design. *Transportation Research Part C: Emerging Technologies*, 107:15–33, 2019. doi:10.1016/j.trc.2019.08.006.

Biyu Wang, Sergio Arturo Ordonez Medina, and Pieter Fourie. Simulation of autonomous transit on demand for fleet size and deployment strategy optimization. *Procedia Computer Science*, 130:797–802, 2018. doi:10.1016/j.procs.2018.04.138.

Jian Wen, Yu Xin Chen, Neema Nassir, and Jinhua Zhao. Transit-oriented autonomous vehicle operation with integrated demand-supply interaction. *Transportation Research Part C: Emerging Technologies*, 97:216–234, 2018. ISSN 0968-090X. doi:10.1016/j.trc.2018.10.018. URL https://www.sciencedirect.com/science/article/pii/S0968090X18300378.

Jizhe Xia, Kevin M. Curtin, Weihong Li, and Yonglong Zhao. A new model for a carpool matching service. *PLOS ONE*, 10(6):1–23, 06 2015. doi:10.1371/journal.pone.0129257. URL https://doi.org/10.1371/journal.pone.0129257.

Felix Zwick, Nico Kuehnel, Rolf Moeckel, and Kay W. Axhausen. Ride-pooling efficiency in large, medium-sized and small towns -simulation assessment in the munich metropolitan region. *Procedia Computer Science*, 184:662–667, 2021. ISSN 1877-0509. doi:10.1016/j.procs.2021.03.083. The 12th International Conference on Ambient Systems, Networks and Technologies (ANT) / The 4th International Conference on Emerging Data and Industry 4.0 (EDI40) / Affiliated Workshops.

Felix Zwick, Gabriel Wilkes, Roman Engelhardt, Steffen Axer, Florian Dandl, Hannes Rewald, Nadine Kostorz, Eva Fraedrich, Martin Kagerbauer, and Kay W. Axhausen. Mode choice and ride-pooling simulation: A comparison of mobitopp, fleetpy, and matsim. *Procedia Computer Science*, 201:608–613, 2022. ISSN 1877-0509. doi:10.1016/j.procs.2022.03.079. URL `https://www.sciencedirect.com/science/article/pii/S1877050922004926`. The 13th International Conference on Ambient Systems, Networks and Technologies (ANT) / The 5th International Conference on Emerging Data and Industry 4.0 (EDI40).

# Appendix A

# Adapting Shared Autonomous Vehicle Strategies: A Multi-City Analysis of Urban Heterogeneity and Public Transport Competition

**Heran Zhao**
*TU Delft*
*Delft, The Netherlands*
*Faculty of Civil Engineering and Geosciences*

**Abstract**

Shared Autonomous Vehicles (SAVs) hold significant potential to redefine urban mobility services. However, the applicability and optimization of their operational strategies in diverse urban contexts remain unclear, particularly concerning the complex interactions with existing public transport systems. This study systematically investigates how urban heterogeneity, including city scale, network topology, and demand patterns, modulates SAV fleet management strategies to balance operator costs and user utility within a multi-modal transportation context.

To this end, this study develops a comprehensive simulation framework that combines an advanced ride-pooling candidate generation algorithm (ExMAS) with an innovative two-stage vehicle assignment optimization process, and integrates a nested Logit model to quantify the competitive dynamics with public transport. This framework is applied to 37 Dutch cities of varying scales to derive generalizable findings.

The study finds that population scale is a fundamental determinant of the required fleet size, exhibiting a strong linear relationship, especially in large cities. The key to enhancing operational efficiency, however, lies in more nuanced urban structure metrics. In large cities, longer commuting distances combined with more complex networks (higher node degrees) foster ride-pooling potential. In contrast, for small cities, the local compactness of the network (higher clustering coefficient) and demand density are critical for improving vehicle turnover efficiency. In competition with public transport, the study's *Public Transport Competitiveness Index* reveals a non-monotonic relationship with travel distance. A state of competitive balance is observed for short-distance trips, while PT holds a distinct advantage in the medium-distance range. For long-distance trips, the inherent speed advantage of SAVs allows them to become the more competitive option.

The findings of this study provide critical strategic insights for SAV operators and policymakers, emphasizing the critical importance of adjusting fleet management strategies according to specific urban characteristics and the competitive environment, thereby providing decision support for achieving an efficient and sustainable urban transportation system.

**Keywords:** Shared Autonomous Vehicles, Fleet Management Strategies, Multi-City Analysis, Ride-Pooling, Multi-Modal Transportation

## I. Introduction

The introduction of Shared Autonomous Vehicles (SAVs) is anticipated to significantly reshape urban transportation. Research indicates that SAVs can mitigate congestion and enhance urban compactness, leading to substantial improvements in regional accessibility and reductions in travel time (Liu et al., 2024). A case study in Tokyo demonstrated a notable mode shift towards SAVs, although it also highlighted the risk of reducing active transport modes like walking and cycling (Ishibashi and Akiyama, 2022). Crucially, the interaction between SAVs and existing public transportation (PT) systems is a key consideration. Cats et al. (2022) found that ride-hailing services exhibit a dual relationship with PT, complementing it by filling service gaps in underserved areas while competing for demand in well-served regions. This complex dynamic underscores the importance of carefully designing operational strategies for SAVs.

Ride-pooling is a core mechanism for improving the efficiency and sustainability of SAV systems. Numerous studies have shown its potential to reduce the required fleet size and total vehicle kilometers traveled (VKT), thereby alleviating congestion and emissions (Fagnant, 2015; Balac et al., 2020; Alonso-Mora et al., 2017). Effective ride-pooling improves vehicle utilization, lowers operating costs, and can reduce passenger waiting times (Jin et al., 2021; Militão and Tirachini, 2021). For instance, Balac et al. (2020) found that ride-pooling could drastically reduce vehicle requirements, while Kucharski and Cats (2020) explored its profitability, noting that fare discounts between 10% and

30% are often necessary to balance operational savings with user incentives. These benefits contribute to the overall sustainability of urban transport by lowering energy consumption and carbon emissions (Fagnant, 2018).

However, the core challenge in deploying SAVs stems from a lack of practical operational experience (Narayanan et al., 2020), making it difficult to predict performance across diverse urban contexts. The effectiveness of ride-pooling is highly dependent on factors such as urban structure, demand patterns, and specific operational strategies (Fagnant, 2018; Soza-Parra et al., 2024; Alonso-Mora et al., 2017). Existing research confirms that urban form and travel demand patterns significantly influence system efficiency. Factors like travel density, the configuration of attraction centers, trip length distribution, and overall demand levels are decisive for SAV performance (Zwick et al., 2021; Soza-Parra et al., 2024; Boesch et al., 2016). For example, Boesch et al. (2016) found that service performance drops significantly at low demand levels, and that acceptable passenger waiting time is a key factor in determining fleet size. In scenarios with low ride-pooling efficiency, the overall benefits of SAVs may not meet expectations (Kumakoshi et al., 2021; Jin et al., 2021). This highlights a critical issue: strategies effective in one urban context may not be transferable to another.

Further research has delved into specific operational strategies. Studies have explored fleet configuration and route optimization (Wang et al., 2018), the optimization of fleet size under different service modes (Monteiro et al., 2021), and the impact of various service level parameters, such as rejection policies and vehicle capacity (Militão and Tirachini, 2021). Other work has focused on integrating external costs like pollution and noise into operational decisions to enhance social benefits (Schröder and Kaspi, 2024). Regarding the integration with existing transport systems, some studies have analyzed the effect of SAVs on traffic congestion in mixed-traffic scenarios (Jin et al., 2021), while others have examined how different combinations of modes, such as SAVs and bicycles, could meet urban travel needs (Fan et al., 2023). Despite these efforts, some analyses have pointed out that the introduction of SAVs could potentially increase total VKT, especially in public transport-oriented cities (Fielbaum and Pudāne, 2024; Fagnant, 2015, 2018), reinforcing the need for integrated planning.

Despite significant progress, key gaps remain in the literature. First, there is a need for a more systematic analysis of how urban heterogeneity affects SAV system performance. Most studies focus on a single city or use simplified indicators, limiting the generalizability of their findings across diverse urban contexts. Second, the integration of SAVs with multimodal transport systems, particularly the competitive dynamics with public transport, has not been sufficiently explored. Existing models often lack a detailed mode choice mechanism, leaving the optimization of the overall transport system unclear. Finally, there is often a disconnect between theoretical ride-pooling schemes and executable, conflict-free vehicle dispatching in operational models.

To address these gaps, this study develops a comprehensive framework to explore how demand characteristics, transport network structure, and competition with public transit jointly affect the performance of SAV systems across different city scales. This research aims to answer the question of:

*In cities of varying scales, how should SAV fleet management strategies be adapted to effectively balance operator and user costs, while considering their competitive interactions with public transport services?*

To address this question within a well-defined scope, this study models a multi-modal transport system where travelers choose between public transport and an on-demand SAV service, which in turn offers both private and shared rides. The analysis adopts a static, offline perspective, evaluating system performance over a complete 24-hour demand cycle. This strategic approach intentionally abstracts away from real-time operational complexities, allowing for a focused and repeatable comparison of fleet management strategies under different urban conditions.

The main contributions of this paper are threefold:

- **Systematic analysis of multi-city heterogeneity**: This study utilizes real-world data from 37 Dutch cities to systematically compare the applicability and optimization pathways of SAV fleet management strategies under different city scales, network structures, and demand patterns.
- **Explicit modeling of competition with public transport**: By integrating a multimodal travel choice model, this study explicitly characterizes the market share and interaction between SAVs and public transport, providing a foundation for their joint optimization.
- **Integrated framework for ride-pooling and dispatch**: This study develops a two-stage approach that combines the generation of high-attractiveness ride-pooling options with an optimization process that produces conflict-free, executable vehicle schedules, bridging the gap between theoretical analysis and practical operations.

The remainder of this paper is structured as follows. Section II. provides a review of the relevant methodologies. Section III. details the simulation framework and the proposed methodology. Section IV. describes the simulation setup, including the case study cities and key performance indicators. Section V. presents and discusses the simulation

results. Finally, Section VI. concludes the paper and offers directions for future research.

## II. Literature Review

Methodologies for SAV fleet management are primarily divided into simulation-based approaches and mathematical modeling. Simulation frameworks are adept at capturing the dynamic and stochastic nature of transport systems. The MATSim (multi-agent transport simulation) framework, in particular, has been widely adopted (Horni et al., 2016). Researchers have extended it with dynamic vehicle routing problem (DVRP) modules (Vosooghi et al., 2019) or demand-responsive transport (DRT) modules (Wang et al., 2018) to optimize fleet operations. Other agent-based tools, such as FleetPy, have been developed to incorporate external costs into routing decisions (Schröder and Kaspi, 2024; Engelhardt et al., 2022). These various tools exhibit a degree of specialization, making them suitable for different research scales and objectives, from real-time operations to long-term network impacts (Zwick et al., 2022; Boesch et al., 2016).

For problems requiring precise optimization, mathematical modeling methods offer distinct advantages. A range of models has been proposed, including multi-objective optimization frameworks (Seo and Asakura, 2022), mixed-integer linear programming (MILP) to determine fleet size (Monteiro et al., 2021; Balac et al., 2020; Militão and Tirachini, 2021), and more complex nonlinear models that account for congestion effects and mode choice (Fan et al., 2023). Other approaches have drawn from graph theory (Qu et al., 2022) or employed compartmental models to describe system dynamics (Jin et al., 2021). However, these methodologies present a trade-off. While simulations capture system complexity, they often lack optimality in specific operational aspects like ride-pooling matching. Conversely, mathematical models provide optimization but can oversimplify real-world conditions or become computationally infeasible for large-scale, dynamic systems.

A central challenge in SAV operations is the efficient matching of shared rides (Xia et al., 2015; Agatz et al., 2012, 2011; Furuhata et al., 2013). To address this, the ExMAS (Exact Matching of Attractive Shared-rides) algorithm was developed as a utility-based method to systematically generate all attractive ride-pooling combinations from a static set of requests (Kucharski and Cats, 2020). Unlike heuristic or real-time methods (Alonso-Mora et al., 2017), ExMAS is designed for strategic analysis, enabling a thorough evaluation of how operational parameters like fare discounts influence pooling potential. This makes it highly suitable for assessing fleet strategies under different urban conditions.

However, while ExMAS provides a powerful foundation for generating high-quality ride-pooling options, its core functionality does not extend to creating conflict-free, executable vehicle schedules. The algorithm excels at assessing theoretical pooling potential but does not solve the operational vehicle dispatching problem. This limitation highlights a critical gap between strategic analysis and practical implementation. This methodological gap, combined with the specific limitations of ExMAS, motivates the integrated framework developed in this study. Table 7 provides a detailed summary of the limitations identified in the reviewed literature.

## III. Methodology

This chapter details the simulation framework developed to analyze SAV fleet operations, as illustrated in Figure 1. The methodology follows a sequential process: it begins with generating travel demand, then creates potential ride-sharing options using the ExMAS algorithm. Subsequently, it models travelers' mode choice between SAVs and public transport. Finally, it employs a novel two-stage optimization for vehicle assignment, which first establishes a feasible decision space of vehicle-to-ride pairings and then applies a global optimization to determine the final vehicle schedules.

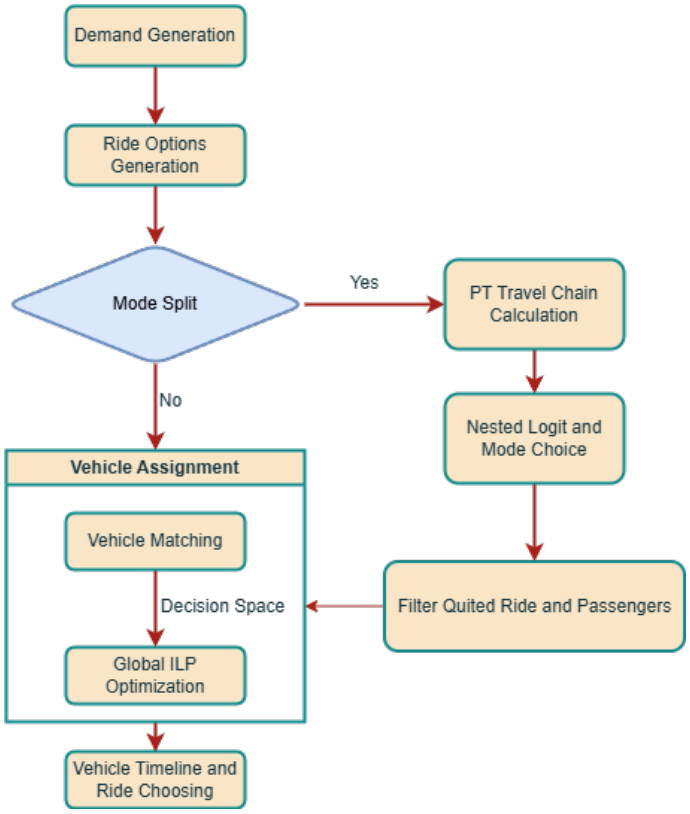



FIGURE 1: SIMULATION FRAMEWORK

## A. *Demand Generation*

The simulation framework is underpinned by a procedural demand generation model designed to create synthetic yet realistic travel requests. This model translates macroscopic travel characteristics into a set of individual trips with specified origins, destinations, and departure times. The methodology consists of two main stages: parameter calibration and spatiotemporal trip generation.

First, the model is calibrated using parameters derived from a suitable external data source, typically a comprehensive travel survey. This calibration ensures that the generated demand reflects observed travel behavior. The key parameters to be calibrated include: the temporal distribution of trips over a 24-hour period, the share of trips by purpose (e.g., commuting versus non-commuting), and the probability distributions of trip lengths, which can be defined separately for each trip purpose.

Second, the trip generation algorithm populates a given road network with individual travel requests. The network nodes are first classified based on land-use data into distinct types, such as *residential*, *work*, and *activity* zones. The algorithm then proceeds chronologically through the simulation period. For each time interval, it generates a number of trips according to the calibrated temporal profiles and purpose shares. The spatial allocation of these trips depends on their purpose. To simulate peak-hour traffic, commuting trips are generated with origins sampled from nodes in *residential* zones and destinations from *work* zones during the morning peak, with the pattern reversed for the evening peak. In contrast, origins for non-commuting trips are sampled from nodes across all zone types to represent more varied travel patterns. For any given trip, once an origin is set, a travel distance is drawn from the corresponding calibrated distribution to identify a suitable destination. Finally, a maximum travel distance can be imposed as a filter to exclude unrealistic long-distance trips.

## B. *Integrated Ride-Pooling and Dispatch Framework*

This study develops an integrated simulation framework to bridge the gap between strategic ride-pooling analysis and operational vehicle dispatch. The framework extends a utility-based ride-pooling generation algorithm with a two-stage vehicle assignment process, designed to produce conflict-free, cost-optimized vehicle schedules from a set of predetermined travel demands.

*Generation of Potential Shared Rides*

The candidate generation process begins by treating each individual travel request as a default single-passenger (or solo) ride. To generate shared-ride options, the framework directly utilizes the open-source ExMAS (Exact Matching of Attractive Shared-rides) package (Kucharski and Cats, 2020). ExMAS systematically attempts to combine the individual trips into attractive shared rides of varying sizes, up to the vehicle's capacity. A shared ride is considered a viable candidate only if the utility gain for each participating passenger, typically from fare discounts, outweighs the disutility of service deviations such as detours and delays. The final set of candidates for consideration therefore includes both the original solo rides and all successfully generated shared-ride combinations.

*Two-Stage Vehicle Assignment*

Once the set of candidate rides is finalized, a two-stage process assigns vehicles to execute them. To facilitate this assignment, vehicles are modeled as individual entities. At the start of the simulation, a fleet is initialized, with each vehicle having a dynamic state (e.g., idle, busy, or inactive) and a continuously updated activity timeline. This timeline records the full schedule of assigned tasks, which is essential for ensuring spatiotemporally consistent and conflict-free dispatching.

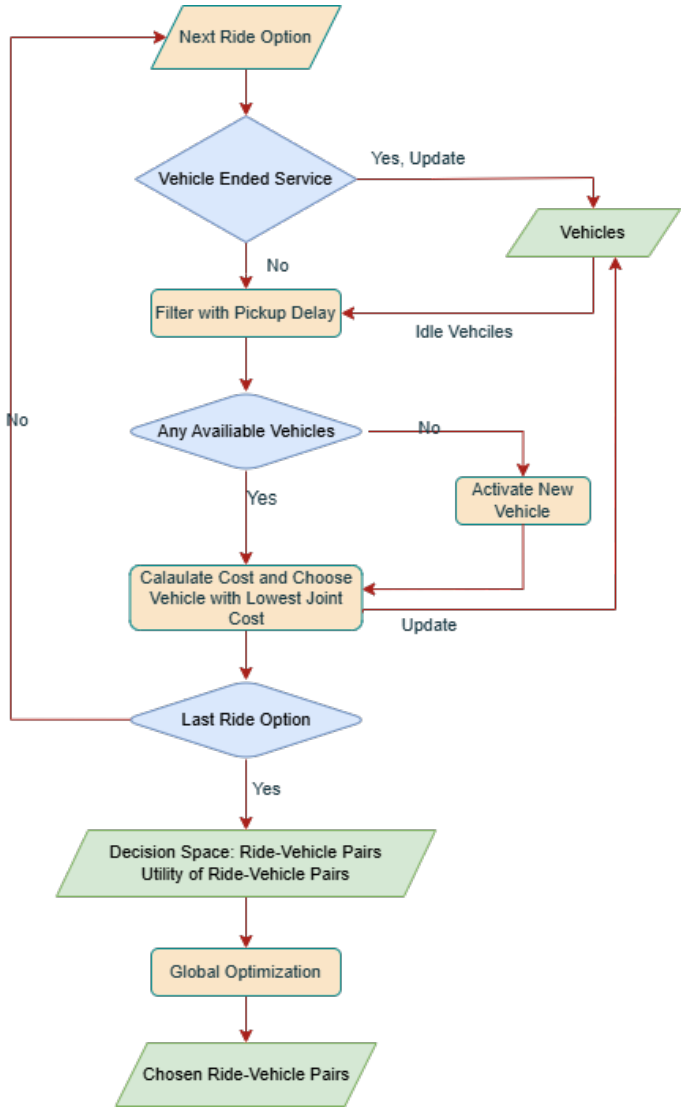



FIGURE 2: VEHICLE ASSIGNMENT FLOWCHART

The first stage generates a preliminary, conflict-free schedule using an event-based simulation that processes rides sequentially, as illustrated in Figure 2. This stage is designed to efficiently reduce the solution space for the final optimization while adhering to key operational constraints. The process begins by sorting all potential rides chronologically by the earliest requested departure time of their constituent passengers. As the algorithm iterates through

each ride, it first dynamically updates the status of all vehicles to ensure that assignments are based on their precise availability and location at that moment.

For each ride under consideration, the core principle is to prioritize the use of existing idle vehicles. The algorithm evaluates every idle vehicle by calculating the travel time from its current location to the ride's first pickup point and the resulting passenger delay. A vehicle is only considered a feasible candidate if this delay is within the passenger's maximum tolerance. This ability to enforce hard constraints is a key feature of this stage. By leveraging the tracked properties of each vehicle entity, the mechanism can implement various operational rules. While this study focuses on a ride-specific service level constraint, the approach is generalizable to other vehicle-related conditions. By pre-filtering candidates based on such feasibility constraints, this stage critically shapes the quality and composition of the solution space for the final optimization. For each feasible vehicle, a joint cost is then calculated, which combines the operator's operational costs (e.g., pickup travel, vehicle waiting) and the passenger's time-based disutility.

A key feature of the framework is its ability to dynamically adapt the fleet size to meet demand, which ensures system robustness by guaranteeing that a feasible assignment option exists for every travel request. If no existing idle vehicle can satisfy a ride's service requirements, the algorithm can activate a new vehicle from a reserve pool. The consequences of this action are governed by a set of pre-defined parameters, including a fixed activation cost and an average waiting time for the passenger, which are incorporated into the joint cost calculation. The number of such activations is also explicitly recorded, providing a key metric for post-analysis.

Finally, to balance solution quality with computational tractability for large-scale scenarios, a greedy assignment strategy is employed. The algorithm compares the joint costs of all feasible ride-vehicle pairs (including those with a potential new vehicle) and provisionally selects the pair with the minimum joint cost. The assigned vehicle's state and timeline are immediately updated. This sequential, greedy approach rapidly produces a complete, conflict-free schedule.

The second stage resolves the final assignment to a global optimum by employing an integer linear programming (ILP) model. This approach adapts the mathematical framework used in the original ExMAS algorithm (Kucharski and Cats, 2020), with a key modification to the objective function. Specifically, this study introduces a system-wide *joint cost function* that combines operator costs and user disutility. This formulation is intentionally chosen to abstract the strategic trade-offs between service quality and operational efficiency. It allows the framework to serve as a tool for evaluating how various operational parameters and policies influence the overall system performance, rather than simply optimizing for a single metric. The ILP model is formulated as a set partitioning problem to select the optimal, conflict-free assignment schedule from the feasible candidates generated in the first stage. The complete optimization model is defined as follows:

The model is formulated as follows:

$$\min \quad \sum_{(r,v)\in \text{Pairs}} c_{r,v} x_{r,v} \tag{1}$$

$$\text{s.t.} \quad \sum_{(r,v)\in \text{Pairs}_p} x_{r,v} = 1 \quad \forall p \in P \tag{2}$$

$$x_{r,v} \in \{0,1\} \quad \forall (r,v) \in \text{Pairs} \tag{3}$$

where $x_{r,v}$ is a binary decision variable that equals 1 if the pre-computed assignment of vehicle $v$ to ride $r$ is selected, and 0 otherwise. The constraint (2) ensures that each passenger $p$ is served exactly once.

The cost of each ride-vehicle pair, $c_{r,v}$, combines the weighted operator cost $O_{r,v}$ and the total user disutility $U_r$:

$$c_{r,v} = \alpha \cdot (f_{bal} \cdot O_{r,v}) + U_r \tag{4}$$

Here, $U_r$ represents the total user disutility for all passengers in ride $r$ (the detailed calculation is provided below), and this component mainly depends on the characteristics of the ride itself, such as passenger composition, route, and timing. $O_{r,v}$ denotes the operator cost incurred by assigning vehicle $v$ to serve ride $r$ (see Section B. for details), which is related to the specific vehicle performing the task (e.g., whether a new vehicle needs to be activated, vehicle waiting time, etc.). To ensure comparability between user utility and operator cost in the optimization process, the operator cost $O_{r,v}$ is first multiplied by an internal balance factor $f_{bal}$, which is automatically calculated based on the relative ranges of the two utility components (as described in Table 14, termed *Balance Factor*). Subsequently, for the purposes of sensitivity analysis and to evaluate different operational strategy preferences, the adjusted operator cost is further multiplied by a manually set weight factor $\alpha$ (corresponding to the *Utility Balance*), allowing the relative

importance of operator cost in the total cost to be tuned during experiments.

The following sections provide a detailed explanation of the calculation methods for user disutility $U_r$ and operator cost $O_{r,v}$.

**Non-Shared Ride Disutility ($U_i^{ns}$)** The disutility of a non-shared ride is composed of the travel fare and the costs associated with time:

$$U_i^{ns} = \underbrace{\pi \frac{l_i}{1000}}_{\text{Fare}} + \underbrace{\beta_{ivt} t_i}_{\substack{\text{In-vehicle} \\ \text{Time Cost}}} + \underbrace{\beta_{wait} \Delta t_i^{wait}}_{\substack{\text{Initial Waiting} \\ \text{Time Cost}}} \tag{5}$$

**Shared Ride Disutility ($U_{i,r}^s$)** For shared rides, the disutility calculation is similar, incorporating a discounted fare and additional time costs related to the shared nature of the trip:

$$U_{i,r}^s = \underbrace{F_{i,r}^s}_{\text{Fare}} + \underbrace{\beta_{ivt} \omega (\hat{t}_{i,r} + |\Delta t_{i,r}^p| + \Delta t^{ba})}_{\substack{\text{Service Process} \\ \text{Time Cost}}} + \underbrace{\beta_{wait} \Delta t_i^{wait}}_{\substack{\text{Initial Waiting} \\ \text{Time Cost}}} \tag{6}$$

Here, $l_i$ represents the direct travel distance in meters, and $\pi$ is the per-kilometer fare rate. For shared rides, a discount $\delta$ is applied. The disutility also includes costs for in-vehicle time $t_i$, initial waiting time $\Delta t_i^{wait}$, and additional time related to the sharing process, such as detours $\hat{t}_{i,r}$, pickup deviations $\Delta t_{i,r}^p$, and boarding/alighting time $\Delta t^{ba}$. The parameter $\omega$ represents the willingness-to-share resistancy, while $\beta_{ivt}$ and $\beta_{wait}$ are the value of in-vehicle and waiting time, respectively.

**Operator Cost** The operator cost $O_{r,v}$ reflects the operational expenses incurred by vehicle $v$ to provide ride $r$:

$$O_{r,v} = \underbrace{c_w t_{r,v}^w}_{\substack{\text{Vehicle} \\ \text{Waiting Cost}}} + \underbrace{(c_f \cdot \mathbb{I}(\text{new}_v))}_{\substack{\text{Fixed Vehicle} \\ \text{Activation Cost}}} + \underbrace{c_t t_{r,v}}_{\substack{\text{Driving} \\ \text{Cost}}} \tag{7}$$

Here, $c_w$ denotes the cost per unit time of vehicle waiting due to early arrival, and $t_{r,v}^w$ represents the total waiting time of vehicle $v$ before starting ride $r$. The term $c_f$ refers to the fixed cost incurred when activating an inactive vehicle, while the indicator function $\mathbb{I}(\text{new}_v)$ equals 1 if vehicle $v$ is newly introduced to the system for this ride, and 0 otherwise. The driving cost per unit time or per kilometer is denoted by $c_t$, and $t_{r,v}$ stands for the total duration or driving time for vehicle $v$ to complete ride $r$, including the travel time to the pickup location.

This comprehensive cost function $c_{r,v}$ allows the optimization framework to balance operational efficiency and user-perceived costs (including time, delay, and fare) when determining the optimal set of ride-vehicle pairs and the required fleet size.

If no existing vehicle can serve a ride within the passenger's maximum delay, a new vehicle is activated. This action incurs a fixed cost ($c_f$). The time required to reach the passenger is represented by a preset parameter, the *Average Waiting Time for New Vehicle* ($t_{wait}^{new}$), which contributes to both the passenger's waiting time disutility and the operator's driving cost. For such activations, the vehicle's own waiting time ($t_{r,v}^w$) is considered zero. This simplified cost model for new vehicles, while necessary, can introduce bias if relied upon excessively. Therefore, the simulation framework is designed to favor solutions that minimize new vehicle activations by ensuring sufficient vehicle availability.

## *C. Integration with Public Transport*

To specifically analyze the competition structure between SAVs and public transport, the core framework is extended with a multimodal choice model. This extension introduces a probabilistic filtering stage that is applied after the ride candidates are generated but before the final vehicle assignment. When this module is active, it first constructs a viable public transport alternative for each travel request and then simulates the passenger's choice, ensuring that the final vehicle assignment stage only considers demand from users who have a propensity for the SAV service.

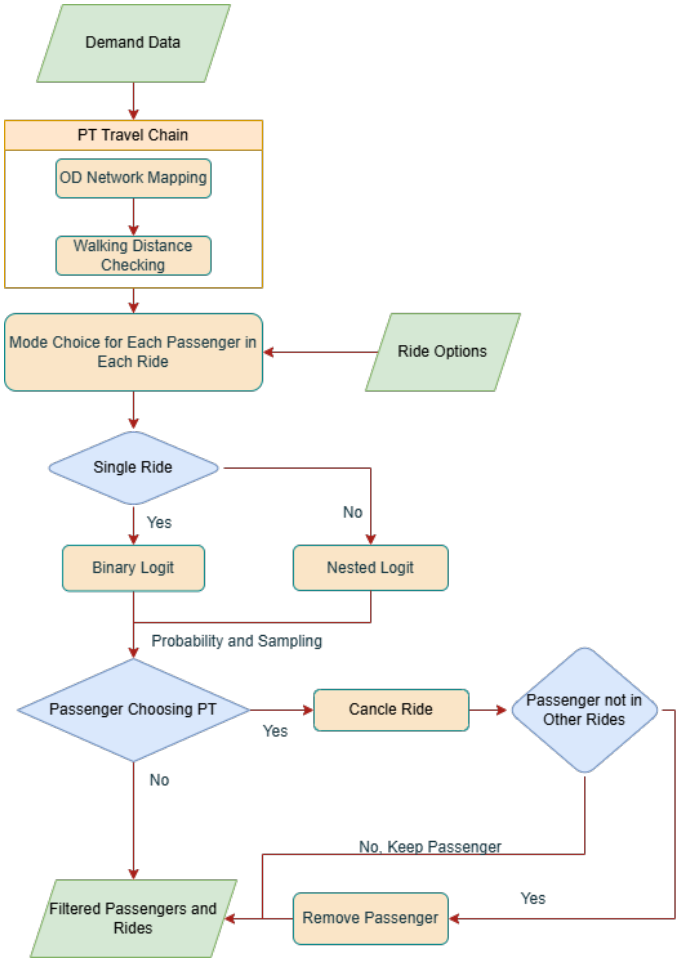



FIGURE 3: PT INTEGRATION FLOWCHART

*1. Modeling the Public Transport Alternative*

A public transport (PT) travel alternative is systematically constructed for each travel request. The process begins by mapping the original trip's origin (O) and destination (D) onto the PT network to create a complete end-to-end travel chain. Specifically, the system identifies the PT station closest to O (station A) and the station closest to D (station B) based on Euclidean distance. The full PT journey is then structured as a three-segment "walk-PT-walk" chain: an initial access walk from O to station A, the main in-vehicle PT leg from station A to station B, and a final egress walk from station B to D. The travel time for each segment is calculated using the corresponding network (walking or PT), which allows for a detailed breakdown of the total travel time into its core components: in-vehicle time and access/egress walking time. For simplification, the model does not explicitly account for station waiting times or transfer penalties; the travel time is calculated based on continuous travel along the determined PT path.

To ensure the realism of the competition, a pre-filtering logic is applied to identify and remove trips for which walking is the most rational choice. This logic first flags a travel request if its constructed PT journey is deemed unreasonable based on two core criteria. The first criterion targets impractically short PT segments: a journey is flagged if its mainline travel distance falls below a predefined threshold. This rule is specifically designed to filter out mapping artifacts where the origin and destination stations are effectively the same point, such as stops on opposite sides of a street. The second criterion flags the journey if its required total walking distance (access and egress) significantly exceeds that of a direct walk. For these flagged requests, a dedicated binary choice model is invoked to compare the utility of the offered SAV service against the utility of a direct walk. If the simulation determines that the passenger chooses to walk, their travel request is removed entirely from the system. This crucial step ensures that the subsequent SAV-PT competition analysis focuses only on trips where both motorized modes represent genuinely competitive options, thereby excluding unrealistically short journeys.

For the remaining trips where PT is a valid alternative, its attractiveness is quantified using a utility function that mirrors the structure used for SAV services, incorporating both monetary and temporal costs. The utility of a PT

trip is formulated as:

$$U_{PT} = (1 - \delta_{PT}) \cdot (p_f^{PT} + p_d^{PT} \cdot d_{PT}) + \beta_{PT,ivt} \cdot t_{PT} + \beta_{walk} \cdot t_{walk}$$

Here, $\delta_{PT}$ is a potential discount, $p_f^{PT}$ and $p_d^{PT}$ are the fixed and distance-based fare components for public transport, $d_{PT}$ is the travel distance, while $t_{PT}$ and $t_{walk}$ represent in-vehicle and walking times, respectively. The parameters $\beta_{PT,ivt}$ and $\beta_{walk}$ quantify the value travelers place on these time components. This calculated utility provides a quantitative basis for the subsequent mode choice simulation.

subsubsectionProbabilistic Mode Choice Filtering With the utility of both SAV and PT alternatives quantified, a probabilistic filtering stage simulates passengers' mode choices using random utility maximization models. The model structure is adapted based on the SAV service type being offered.

For a solo ride offer, a standard binary logit model is used to compute the choice probability between the non-shared SAV ride and the PT alternative. The probability of choosing mode $m$ is given by:

$$P(m) = \frac{e^{V_m}}{\sum_{k \in \{\text{non-shared}, PT\}} e^{V_k}}$$

where $V_m = -U_m$ is the systematic utility (cost) of mode $m$.

For a shared ride offer, a nested logit model is employed to capture the higher correlation between shared and non-shared SAV services. The model has a two-level structure: the upper level models the choice between the 'SAV' nest and the 'PT' alternative, while the lower level models the choice between shared and non-shared services within the SAV nest.

It is critical to clarify the function of this nested model within the framework. Its sole purpose is to determine the probability of a passenger choosing public transport, not to assign them to a specific SAV service based on their preference. If the simulation indicates a passenger chooses the SAV nest, they simply remain in the demand pool for the final optimization. The decision of whether they are ultimately assigned a shared or non-shared ride is made by the global cost minimization algorithm (Equation 1), which considers the entire system's efficiency. The attractiveness of the SAV nest is given by the Logsum term:

$$IV_{SAV} = \lambda_{SAV} \ln \left( e^{V_{\text{shared}}/\lambda_{SAV}} + e^{V_{\text{non-shared}}/\lambda_{SAV}} \right)$$

where $\lambda_{SAV}$ is the scale parameter for the SAV nest ($0 < \lambda_{SAV} \leq 1$), reflecting the correlation among alternatives within the nest. The probability of choosing the PT alternative is then calculated as:

$$P(PT) = \frac{e^{V_{PT}}}{e^{IV_{SAV}} + e^{V_{PT}}}$$

Finally, a stochastic simulation is performed for each passenger based on these choice probabilities. If any passenger within a candidate shared ride is simulated to choose PT, that entire shared ride is invalidated and removed from the candidate set. This filtering process yields a refined demand pool that only includes requests with a clear propensity for SAV service, which then forms the final input for the assignment optimization stage.

## IV. Simulation Setup

This study applies the developed simulation framework to investigate how urban heterogeneity and public transport competition influence SAV system performance across a diverse set of real-world urban contexts. Using travel demand and network data for 37 Dutch cities, the analysis is grounded in a robust baseline scenario established for each city through a detailed calibration process. Building on this foundation, two primary analyses are conducted: a multi-city correlation analysis to identify the key urban drivers of SAV performance, and a scenario-based analysis to explore the competitive dynamics between SAVs and PT under different pricing and congestion conditions.

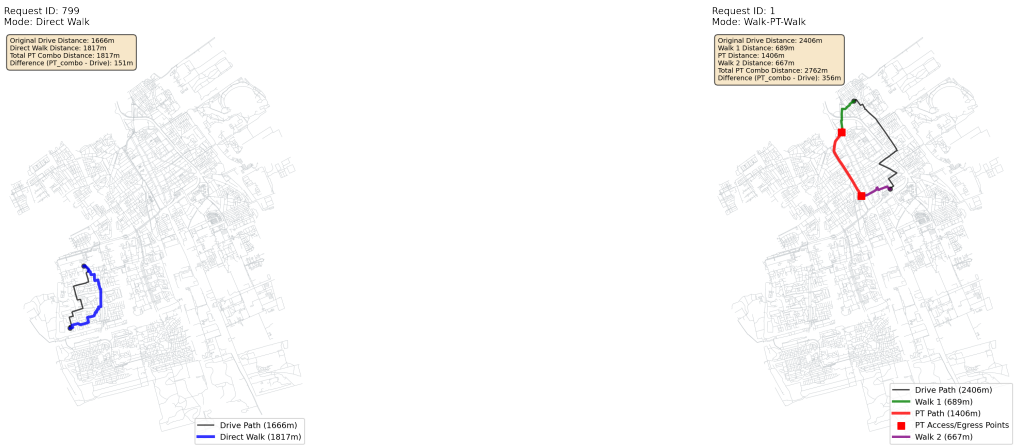
### *A. Case Study and Data Sources*

The case studies for this research were selected from the ODiN dataset, a comprehensive national travel survey for the Netherlands (voor de Statistiek , CBS). To ensure statistical reliability, an initial screening identified all municipalities with more than 1,000 travel records; this process yielded 37 cities of varying scales that were selected for analysis. The dataset provides detailed trip characteristics for each municipality—including travel distance, departure time, and purpose—which form the statistical basis for calibrating the local demand patterns in the simulation. While

ODiN provides the statistical basis for demand patterns, the total number of simulated trips for each city is determined through a scaling procedure. The city's total daily travel demand is first estimated by multiplying its population by a national average trip frequency rate. To ensure computational tractability for the large-scale multi-city analysis, a scale factor of 1% is applied to this total estimated demand. This 1% represents a sample of the total urban travel volume, not the market share of any specific service.

The transport networks for all three modes considered in this study—road, walking, and public transport (PT)—were extracted from OpenStreetMap (OSM) (OpenStreetMap, 2024). The public transport competition analysis considers all available rail-based modes, specifically including network elements tagged in OSM as *tram*, *light rail*, *subway*, or *rail*. Most track data for these PT networks were imported by the official infrastructure manager, ProRail, after 2013, and station data were updated after 2020 (ProRail, 2025).

The walking and public transport networks provide the basis for constructing multimodal travel alternatives. For each request, a "walk-PT-walk" trip chain is generated as the primary PT alternative. A direct walking route is also constructed, serving as a baseline in a pre-filtering step to remove trips where walking is the more rational choice. This ensures the subsequent mode choice analysis focuses on trips where SAV and PT are genuinely competitive. Figure 4 illustrates the outcome of this path generation for both types of routes.




(A) Example of a direct walking route.

(B) Example of a public transport trip chain.

Figure 4: Illustration of route construction from OSM data for a direct walking trip and a public transport alternative.

Figure 5 provides an example of functional zone classification for Amsterdam. To illustrate the outcome of the demand calibration process based on this framework, Figure 6 presents the results for Amsterdam as a representative case. The generated temporal distribution successfully replicates the bimodal commuting pattern observed in the ODiN data, with distinct morning and evening peaks (Figure 6a). Similarly, the distribution of simulated trip distances aligns closely with the empirical data, particularly for short to medium-distance trips that constitute the majority of urban travel (Figure 6b). While minor deviations exist for longer trips—a consequence of random node sampling on the network—the overall generated demand provides a robust and realistic foundation for the subsequent analyses.

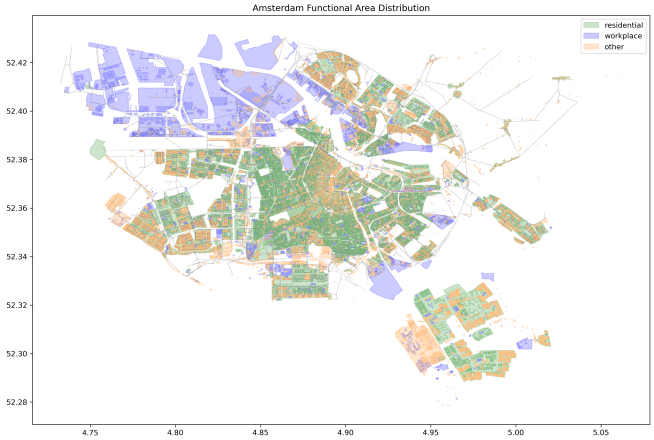
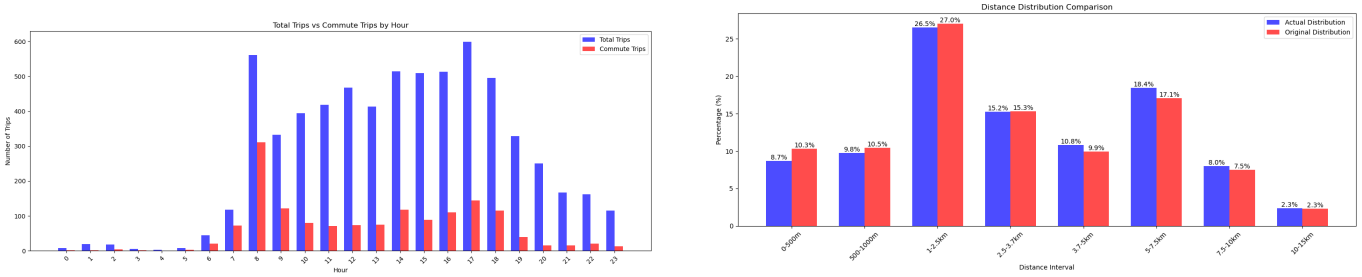



Figure 5: Example of Functional Zone Division and Road Network Node Classification in Amsterdam.




(a) Temporal Distribution

(b) Trip Distance Distribution

Figure 6: Demand generation calibration results for Amsterdam, comparing simulated data with the ODiN dataset.

## B. *Parameter Settings and Calibration*

To ensure the reliability of the analysis, this study establishes a baseline scenario by setting default values for key model parameters. These parameters are categorized into three groups based on their nature and calibration method.

The first group of parameters establishes the fundamental service level assumptions for the baseline scenario. This study assumes a *Maximum Pickup Delay* of 600s and a *Shared Discount* of 30%. To ensure operational consistency within the model, the *Average Waiting Time for New Vehicle* is set to 400s, a value lower than the maximum delay to prevent illogical vehicle dispatching choices. A comprehensive list of all fixed model parameters, such as those related to value-of-time and costs, is provided in Appendix C.

To account for local heterogeneity, another set of key parameters is calibrated individually for each city. The *Willingness to Share Factor* ($\omega$) is adjusted to match the observed ride-pooling rate from the ODiN dataset for each city, ensuring the model's pooling potential reflects real-world user preferences. Similarly, the *Initial Proportion* ($p_{init}$) is calibrated for each city to establish a stable initial fleet size that minimizes the need for new vehicle activations. This city-specific approach ensures that the baseline for each urban environment is appropriately scaled.

In addition, key operational parameters governing the system's economic behavior are calibrated to a single set of default values for all experiments to ensure a consistent and comparable baseline. The *Waiting Cost* ($c_w$) is set to 0.05 €/s. This value was determined through a sensitivity analysis showing that it effectively penalizes excessive vehicle waiting by making its cost comparable to driving costs, without causing significant negative impacts on pickup efficiency or passenger wait times. The *Utility Balance* ($\alpha$), which weights operator costs relative to user disutility in the optimization function, is set to 0.06. This value was identified as a point of diminishing returns where further emphasis on operator costs yields smaller efficiency gains while continuing to degrade the user experience, thus representing a strategic balance between operational efficiency and service quality for the baseline scenario.

### C. Key Performance Indicators

System performance is evaluated using a series of key performance indicators (KPIs). These metrics are grouped into several thematic categories: fleet size and utilization, system-wide operational efficiency, user-perceived service quality, and mode choice dynamics. A complete list of all KPIs, with their precise definitions and formulas, is provided in Appendix B.

### D. Simulation Scenarios

**Multi-City Correlation Experiment** This analysis utilizes the calibrated baseline scenario across all 37 case study cities to systematically investigate the relationship between urban heterogeneity and SAV system performance. To do so, a comprehensive framework of indicators was established to characterize cities across three key dimensions: population scale, network structure, and demand patterns; the detailed definitions for these indicators are provided in Appendix D. The analysis is conducted in several stages. First, a qualitative group comparison is performed to identify high-level trends in KPIs across large, medium, and small city categories. This is followed by a detailed quantitative correlation analysis, examining the statistical relationships between the urban indicators and KPIs, both for the entire set of 37 cities and within each size group to isolate scale-dependent effects. The objective of this multi-stage analysis is to identify robust relationships and provide generalizable insights into how urban context shapes SAV fleet requirements and operational efficiency.

**PT Competition Experiment** To investigate the competitive dynamics between SAVs and public transport, this study designs experiments focusing on two key factors: PT pricing strategies and traffic congestion. These are systematically varied using the model parameters *PT discount* and SAV *avg speed*, respectively.

The analysis consists of a primary experiment to assess the impact of PT pricing and a supplementary grid search to explore the joint effects of pricing and congestion. The parameters for these experiments are detailed in Table 1 and Table 2.

Table 1: Scenarios for PT Discount Sensitivity Analysis

| **Parameter** | **Value(s)** |
|---|---|
| *PT discount* | 10%, 20%, ..., 90% |
| *avg speed* | 8 m/s (Baseline) |

Table 2: Scenarios for Joint Effects

| **Parameter** | **Value(s)** |
|---|---|
| *PT discount* | 0%, 10%, ..., 90% |
| *avg speed* | 5, 6, 7, 8, 9, 10 m/s |

The impact of traffic congestion is analyzed by systematically varying the SAV *avg speed* as a proxy for different congestion levels. This analysis is conducted for the morning peak hour (8:00-9:00), when congestion effects are most pronounced, and the PT speed is held constant throughout the scenarios.

## V. Results

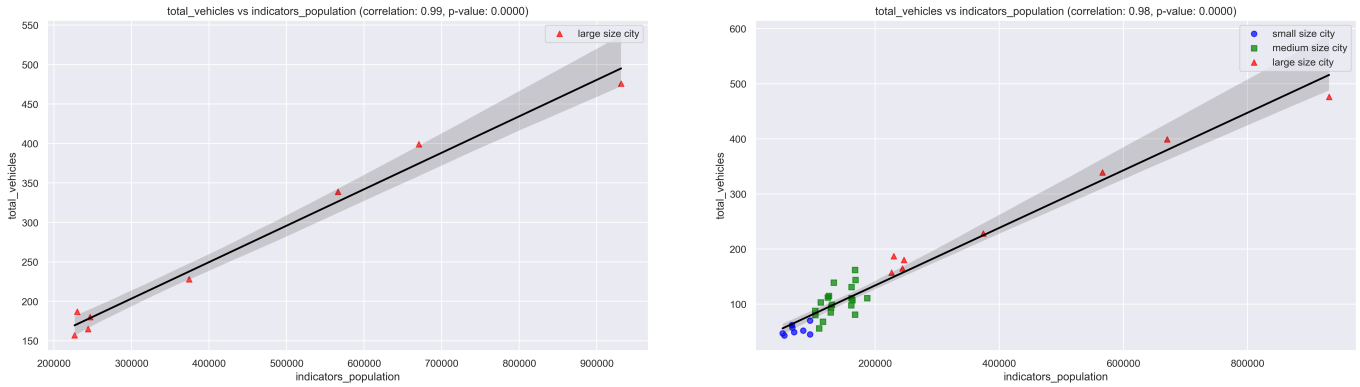
### A. Multi-City Correlation Analysis

This section analyzes the relationship between urban characteristics and SAV system performance. The analysis reveals that fleet size is primarily driven by population scale, while operational efficiency is more closely linked to the nuances of urban network structure, with different factors being dominant in cities of different sizes.

The most direct determinant of the required fleet size is population. Across all 37 cities, *population* exhibits an exceptionally strong positive correlation with both *total vehicles* (r=0.976) and *peak concurrent vehicles* (r=0.980), as detailed in Table 3. This near-linear relationship, visualized in Figure 7, is most pronounced in large cities (r=0.991), confirming that at a macro level, resource requirements scale directly with city size.

Table 3: Key Correlations between City Indicators and KPIs across all 37 Cities.

| City Indicator | KPI | Correlation | P-Value |
|---|---|---|---|
| population | total vehicles | 0.9758 | <0.001 |
| population | peak concurrent vehicles | 0.9796 | <0.001 |
| average degree | saved miles ratio | 0.5906 | <0.001 |
| road density | average pickup time | -0.5584 | <0.001 |
| population density | average pickup time | -0.5111 | 0.002 |




(a) Large Cities (N=8)

(b) All Cities (N=37)

Figure 7: Relationship between Total Vehicles and Population.

While population dictates the necessary scale of the fleet, operational efficiency is driven by the underlying urban structure, with different mechanisms at play in large and small cities.

In **large cities**, ride-pooling efficiency is strongly associated with trip length and network connectivity. The *pooling ratio* is highly correlated with both a longer *average commuting distance* (r=0.884) and a higher *average degree* of the road network (r=0.807). This suggests a mechanism where longer trips provide greater incentives for sharing, while a more connected network offers more routing options to efficiently combine them.

In contrast, the factors influencing performance in **small cities** are more complex. On one hand, vehicle utilization is enhanced by local network topology and density. A higher *average cluster coefficient*, indicating a tightly-knit local road network, is strongly correlated with a better *vehicle reuse rate* (r=0.839). Similarly, higher *population density* is associated with more *average rides per vehicle* (r=0.750). On the other hand, a longer *average commuting distance* is positively correlated with the required *total vehicles* (r=0.802). Unlike in large cities where it promotes pooling, in smaller cities, longer trips increase vehicle occupancy time, reducing turnover and thus demanding a larger fleet to maintain service levels.

The influence of network infrastructure on operational efficiency is most clearly observed in **medium-sized cities**. In this group, higher *road density* shows a significant negative correlation with *average pickup time* (r=-0.540). This effect is less pronounced in other city scales; large cities tend to have uniformly high density, which reduces statistical variance, while data from small cities may exhibit more noise. Medium-sized cities, therefore, provide a distinct context where a well-developed road network directly facilitates more efficient vehicle dispatching.

Table 4: Selected Scale-Specific Correlations between City Indicators and KPIs.

| City Scale | City Indicator | KPI | Correlation | P-Value |
|---|---|---|---|---|
| Large (N=8) | avg commuting distance | pooling ratio | 0.884 | 0.004 |
| Large (N=8) | average degree | pooling ratio | 0.807 | 0.016 |
| Medium (N=19) | road density | average pickup time | -0.540 | 0.017 |
| Small (N=8) | avg commuting distance | total vehicles | 0.802 | 0.017 |
| Small (N=8) | average cluster coefficient | vehicle reuse rate | 0.839 | 0.009 |
| Small (N=8) | population density | avg rides per vehicle | 0.750 | 0.032 |

Service quality, specifically passenger pickup time, is primarily determined by urban density. Across the entire sample, both higher *road density* (r=-0.558) and *population density* (r=-0.511) are correlated with lower average pickup times (see Table 3). The effect of road network density is particularly evident in medium-sized cities. However, this relationship weakens in large cities, likely due to a lower variance in road density among them, and becomes statistically insignificant in small cities, where a small sample size combined with high variance in density may obscure the underlying trend. This indicates that while density is a key driver of service quality, its statistical detectability varies with city scale and data characteristics.

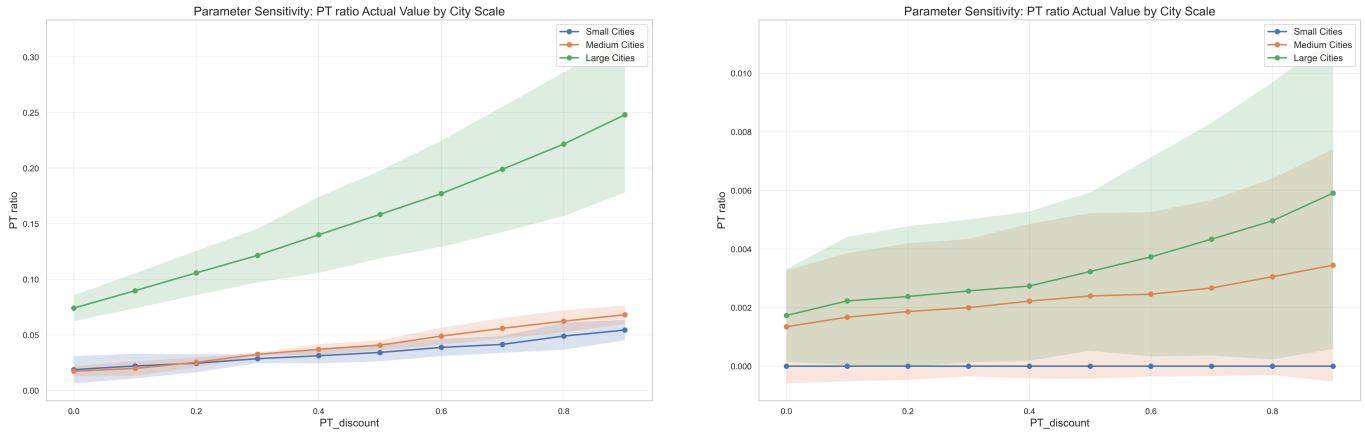
## *B. PT Competition Analysis*

The case study cities are categorized based on their baseline scenario PT mode share from a preliminary simulation. Cities with a PT ratio greater than 1% are grouped into the *Effective Competition Set* (detailed in Table 5), as this level of usage indicates foundational network competitiveness. The remaining cities form the *Network-Restricted Set*.

Table 5: Cities Constituting the 'Effective Competition Set' (PT Ratio > 1%)

| City | PT Ratio |
|---|---|
| Rotterdam | 19.77% |
| 's-Gravenhage | 15.85% |
| Amsterdam | 11.89% |
| Delft | 4.39% |
| Houten | 4.33% |
| Zoetermeer | 3.75% |
| Nieuwegein | 3.00% |
| Amstelveen | 2.92% |

In the *Effective Competition Set*, the PT mode share responds sensitively to increasing discounts, demonstrating that pricing can be an effective competitive tool when a functional network is in place. Conversely, in the *Network-Restricted Set*, the mode share remains largely unresponsive in small cities. This outcome confirms that in these cities, the primary constraint on PT competitiveness is the physical network, not the price.




(a) PT mode share (Effective Competition Set) (b) PT mode share (Network-Restricted Set)

Figure 8: Impact of PT Discounts on PT Mode Share across City Groups.

The analysis is extended to consider the role of traffic congestion, modeled by systematically varying the SAV average speed. Figure 9 reveals a synergistic effect between congestion and pricing on mode choice. As congestion worsens (i.e., SAV speed decreases), the PT mode share becomes increasingly sensitive to price incentives. The reduced SAV speed increases its travel time, making the PT alternative more attractive and thus amplifying the impact of discounts. This finding reinforces that the effectiveness of PT pricing is highly context-dependent, relying not only on the intrinsic quality of the PT network but also on the performance of competing modes.

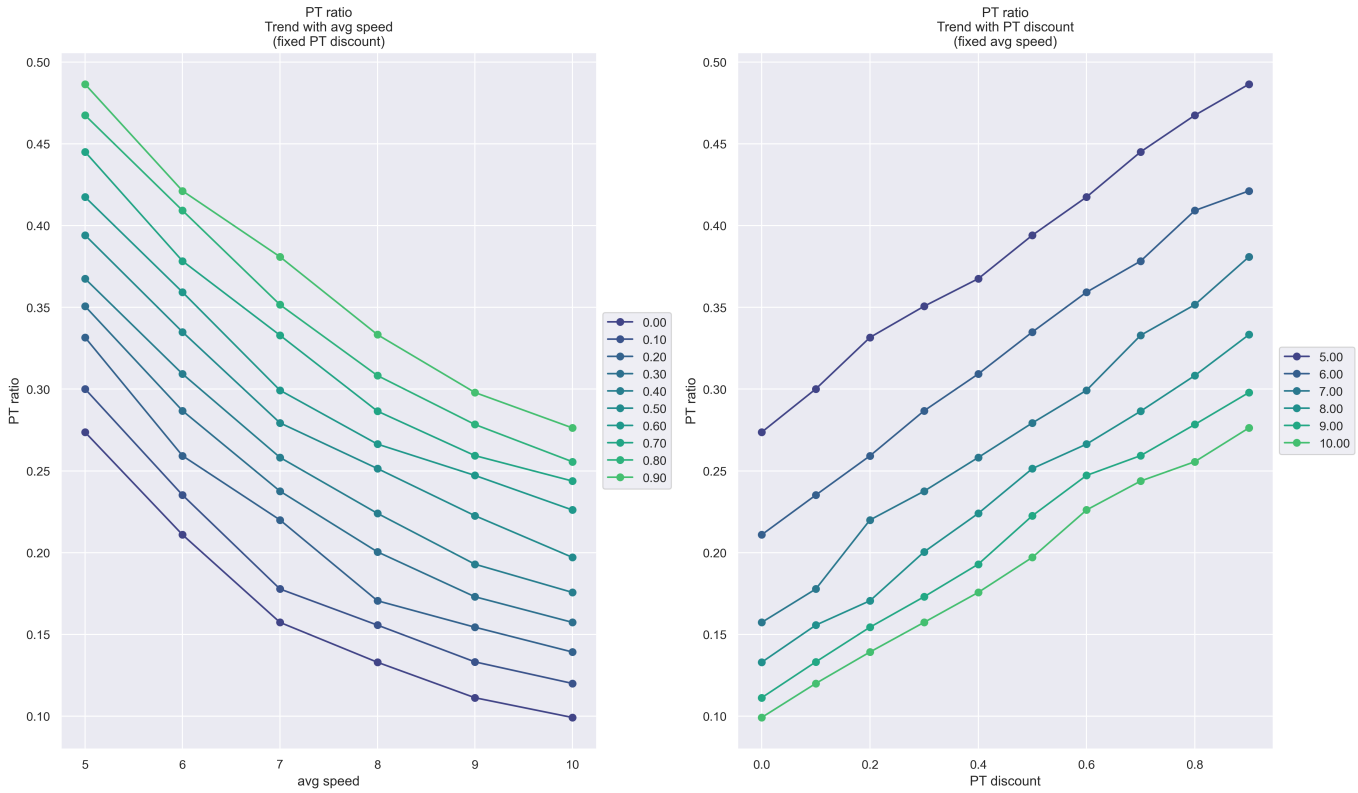
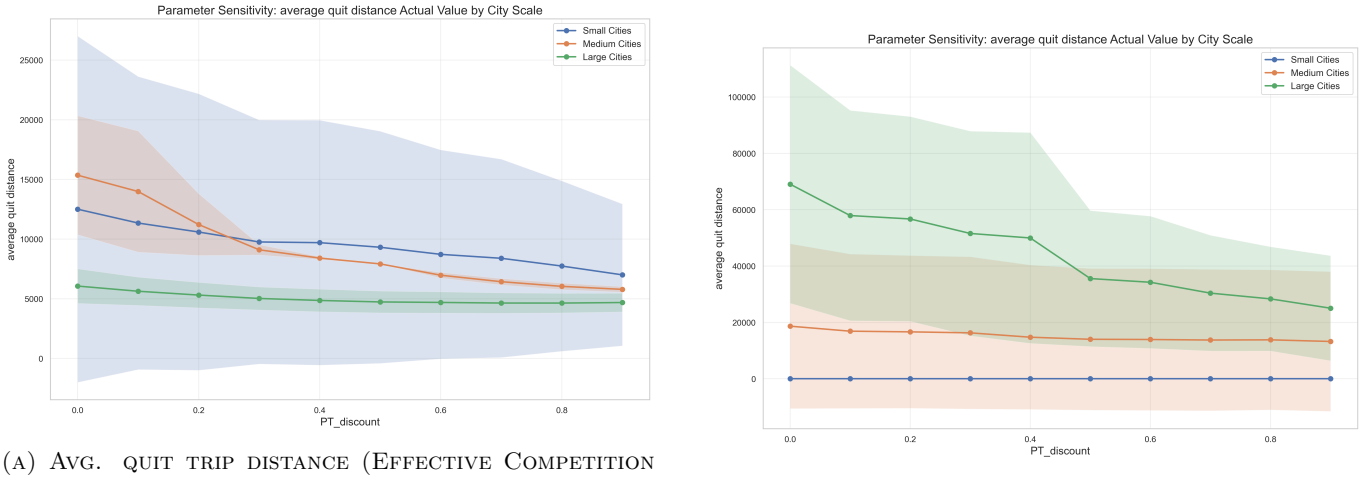



Figure 9: Interaction Effect of Average SAV Speed and PT Discount ($\delta_{PT}$) on PT Mode Share.

Further analysis reveals that the competitive dynamics also vary by trip distance, as shown in Figure 10. In the *Network-Restricted Set*, PT competition is predominantly limited to long-distance trips due to the network's focus on trunk corridors, resulting in a higher average quit distance. In contrast, within the *Effective Competition Set*, PT competes across a broader range of distances. As discounts increase, this group attracts users with shorter trip distances, leading to a reduction in the average quit distance.




(a) Avg. quit trip distance (Effective Competition Set)

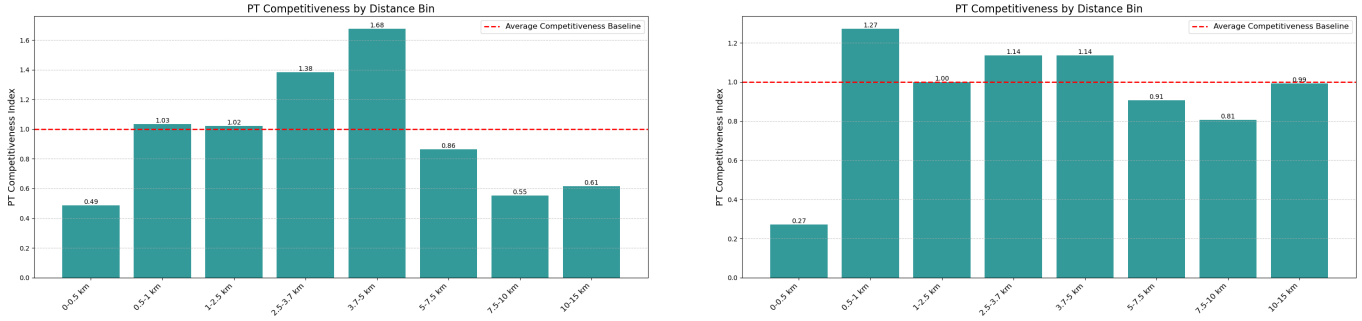
(b) Avg. quit trip distance (Network-Restricted Set)

Figure 10: Average Original Trip Distance of Users Switching to PT at Different Discount Levels.

To investigate why PT discounts are particularly effective at attracting shorter-distance travelers, the analysis introduces a *Public Transport Competitiveness Index* ($CI_i$). For a specific distance bin $i$, this index is defined as the ratio of the proportion of trips switching to PT within that bin to the proportion of total trips in the same bin:

$$CI_i = \frac{P_i}{D_i} = \frac{n_{i,PT}/N_{PT}}{n_i/N_{OD}} \tag{8}$$

where $n_i$ represents the number of original trips in bin $i$, $n_{i,PT}$ denotes the number of trips switching to PT in that bin, and $N_{OD}$ and $N_{PT}$ are the respective totals of original and PT-switching trips. A value of $CI_i > 1$ indicates a competitive advantage for PT in that distance segment. This index can be further expressed as the ratio of the switch rate within the bin ($R_i = n_{i,PT}/n_i$) to the global average switch rate ($R_{avg} = N_{PT}/N_{OD}$), simplifying to $CI_i = R_i/R_{avg}$.




(A) Baseline CI ($\delta_{PT} = 0$). (B) CI with Discount ($\delta_{PT} = 0.5$).

Figure 11: PT Competitiveness Index by Distance for Rotterdam.

The case of Rotterdam, presented in Figure 11a, illustrates the underlying mechanism. The analysis reveals a non-monotonic relationship between PT competitiveness and travel distance. A competitive balance ($CI_i \approx 1$) is observed for trips between 0.5 and 2.5 km, while PT demonstrates a clear advantage in the medium-distance range of 2.5 to 5 km, peaking at a competitiveness index of 1.68. For trips longer than 5 km, SAVs regain a competitive edge, likely due to their speed advantage over longer distances. This observed competitive structure suggests that the effectiveness of PT discounts is not uniform but is instead dependent on the intrinsic competitiveness of the PT network at different travel distances.

The impact of this pricing strategy extends to the travel patterns of those who switch to PT, as illustrated in Figure 12. While higher PT discounts reduce the average total trip distance of switchers (as seen in Figure 10), they simultaneously increase the average distance of the PT leg for these same trips. This effect is a direct consequence of the competitive dynamics revealed by the CI analysis: the discounts are powerful enough to attract a large volume of medium-distance travelers, which lowers the average overall trip length, while at the same time making PT cost-effective enough to successfully compete for a larger share of longer-distance journeys, thus increasing the average length of the PT segment itself.

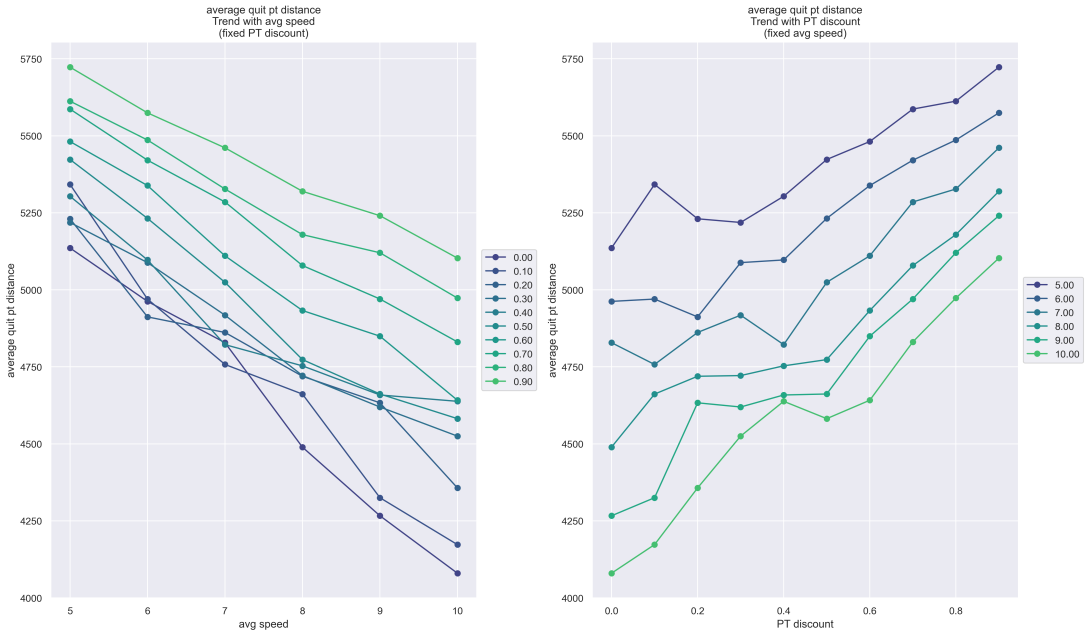
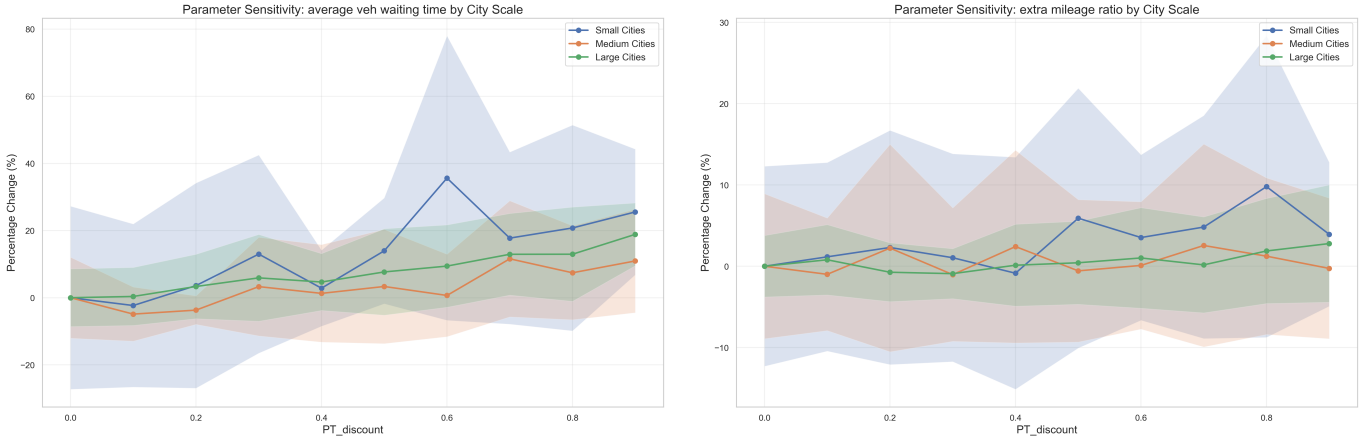



Figure 12: Impact of PT Discount on the Average Distance of the PT Leg for Switching Trips.

The competitive shift in demand directly influences SAV operational efficiency. In the *Effective Competition Set*, increasing PT discounts reduce SAV demand density, leading to a slight rise in the *extra mileage ratio* and *average vehicle waiting time* (Figure 13). Interestingly, the *pooling ratio* of the remaining SAV trips increases (Figure 14). This is because the choice model filters out a higher proportion of less-poolable solo SAV trips, thereby increasing the concentration of trips with greater pooling potential in the remaining demand.




(a) Average Vehicle Waiting Time

(b) Extra Mileage Ratio

Figure 13: Impact of PT Competition on SAV Operational Efficiency (Effective Competition Set).

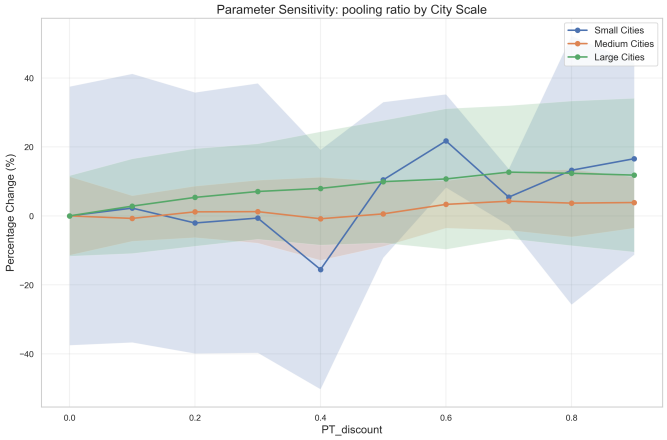
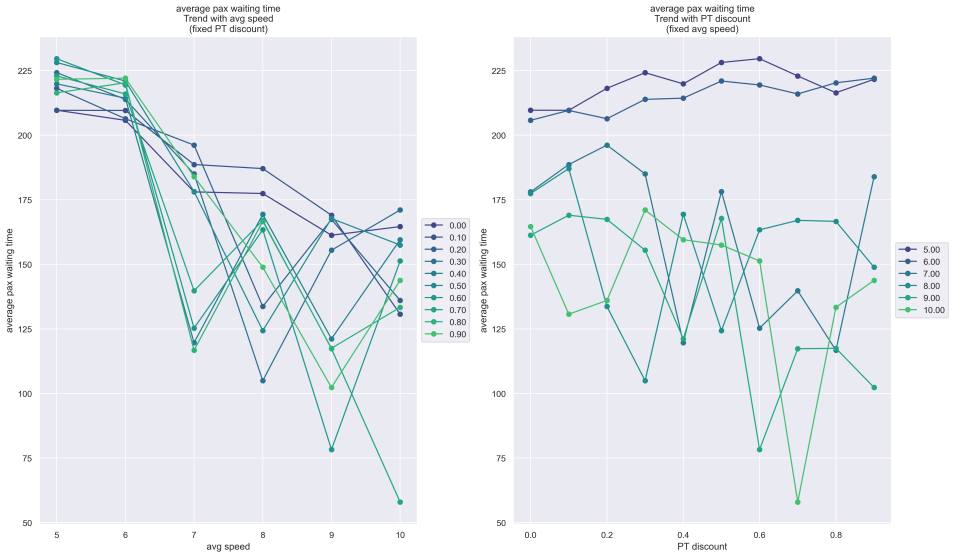
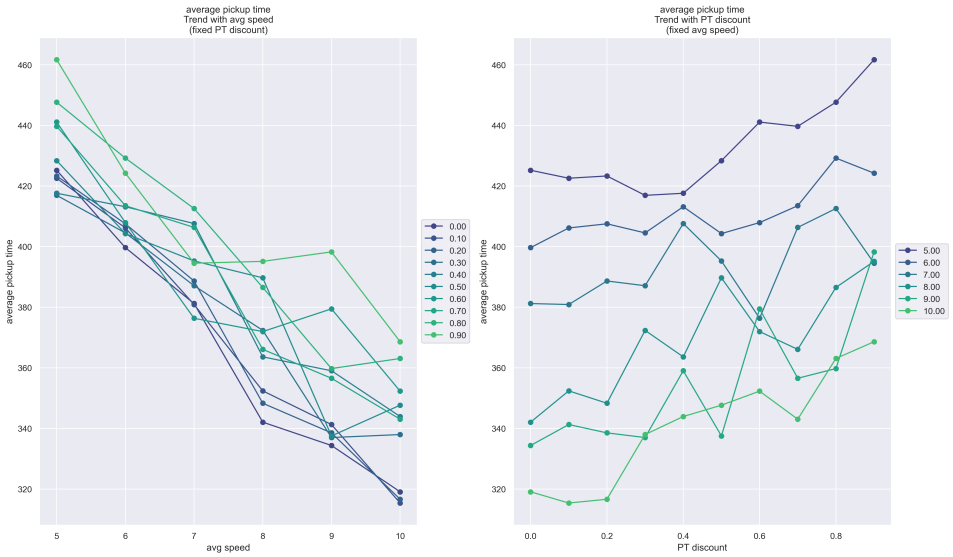



Figure 14: Impact of PT Competition on Pooling Ratio (Effective Competition Set).

Finally, the direct impact of congestion on SAV service metrics is examined. As shown in Figure 15, SAV speed is the dominant factor determining passenger waiting and pickup times. Waiting time exhibits a non-linear response, increasing sharply at very low speeds, while pickup time shows a more direct, linear relationship. Notably, the performance curves across different discount levels remain nearly parallel, indicating a weak interaction. This suggests that congestion acts as a primary physical constraint on the system, while PT pricing primarily influences demand distribution between the modes.




(A) Average Passenger Waiting Time




(B) Average Pickup Time

Figure 15: Impact of SAV Speed and PT Discount ($\delta_{PT}$) on SAV Service Metrics.

In summary, the analysis reveals a clear hierarchy of effects: the underlying PT network determines the potential for competition, pricing incentives activate this potential by reshaping demand across different travel distances, and traffic congestion acts as an overarching physical constraint on the operational performance of the SAV system.

## VI. Conclusion and Discussion

This study developed a comprehensive simulation framework to investigate SAV fleet management across 37 Dutch cities, integrating an advanced ride-pooling algorithm with a two-stage vehicle assignment optimization process. The framework bridges the gap between theoretical ride-pooling potential and executable vehicle scheduling through a filtering mechanism based on discrete-event simulation, followed by integer linear programming to determine the optimal assignment scheme. Additionally, a nested Logit model simulates mode choice between SAVs and public transport (PT), with detailed construction of PT travel chains enabling accurate representation of multimodal travel behavior.

The investigation reveals significant insights into operational efficiency and competitive dynamics. City *scale* emerges as a fundamental determinant of SAV performance, with large cities exhibiting enhanced operational efficiency due to high demand density, evidenced by lower *extra mileage ratio* and reduced *average pickup time*. In contrast,

small cities require higher initial *fleet size* relative to demand due to geographically dispersed travel patterns, resulting in longer vehicle occupancy times and lower turnover efficiency. The study further categorizes cities into an *Effective Competition Set*, where PT mode share exceeds 1% at baseline, and a *Network-Restricted Set*, where PT infrastructure limits competitiveness. Within the *Effective Competition Set*, increasing PT discounts effectively elevate PT mode share, particularly attracting short- and medium-distance travelers, while in the *Network-Restricted Set*, mode share remains largely unresponsive to price incentives due to network constraints. Travel distance emerges as a pivotal factor in this competition, with SAVs holding a structural advantage for short trips due to door-to-door convenience, while PT demonstrates greater competitiveness in medium to long-distance segments (5-15 km). For very long trips, the speed advantage of SAVs often outweighs PT's cost benefits. The competitive dynamics also impact SAV operational metrics, with rising PT discounts reducing SAV demand density, leading to slight increases in *average vehicle waiting time* and *extra mileage ratio*, while the *pooling ratio* among remaining SAV trips rises due to the filtering out of less-poolable solo trips.

For SAV operators, these findings suggest several strategic approaches. Resource allocation should be adjusted based on urban characteristics, with higher initial vehicle provision in small cities to ensure service reliability. In large cities, the operational focus should leverage density advantages, while in small cities, operational efficiency can be improved by lowering the weighting of passenger-related costs. The topology of the road network also influences system performance, with high local clustering in small cities offering greater potential for vehicle reuse. In medium-sized cities, designating high-density areas as core operational zones while implementing specialized dispatch strategies for low-density areas can address efficiency limitations. Regarding competition with PT, operators should recognize the growing competitive pressure in medium to long-distance markets, especially as PT pricing policies are introduced. Strategies such as dynamic pricing, targeted service differentiation, or collaboration with PT may be necessary to retain core user groups. The strategic focus can shift toward dynamically reallocating fleet resources to segments where SAVs have a comparative advantage, such as short-distance trips or areas with insufficient PT coverage.

Despite these insights, several limitations define the scope of the conclusions. The static and offline nature of the model, with ride-pooling generation assuming all travel demands are known in advance while vehicle assignment employs a sequential approach, does not achieve fully optimal vehicle-ride matching. The single-day demand simulation fails to capture temporal *variability* such as weekday-weekend differences or seasonal fluctuations, restricting the assessment of long-term market dynamics. The objective function focuses on cost minimization, excluding revenue models and broader social benefits, limiting the evaluation of economic viability or external impacts. In modeling PT competition, the system is simplified to rail networks, excluding bus services and estimating travel times based on network distance and average speed rather than actual timetables or service frequencies. Operational simplifications include assuming a single, monopolistic operator and achieving vehicle relocation indirectly through availability and waiting time thresholds. Moreover, the input demand represents a 1% sample of total urban travel, potentially missing nonlinear *scale* effects at real-world demand levels, while parameter calibrations such as uniform *average vehicle speed* across cities further limit direct applicability to diverse contexts.

Future research can address these limitations through several targeted directions. Expanding the model from single-city analyses to regional and intercity travel networks would provide insights into comprehensive travel chains, requiring improved PT modeling with actual timetables and service frequencies. Refining operational strategies should explore explicit vehicle relocation mechanisms, reservation-based paradigms, and hybrid market simulation frameworks combining centralized optimization with agent-based models of independent driver behavior. For competitive strategies with PT, research should focus on detailed pricing mechanisms beyond uniform discounts, considering passenger heterogeneity and market segmentation. Constructing multilayer network models for multimodal transport systems and developing specific topological indicators would deepen understanding of how network structure influences mode choice and system efficiency. Lastly, integrating broader socioeconomic impacts into optimization objectives by quantifying externalities like air pollution or noise would enable more holistic assessment of SAV systems, while addressing real-world operational constraints such as charging infrastructure or seasonal demand fluctuations would ensure findings offer practical strategic guidance for sustainable urban mobility.

## References


Agatz, N., Erera, A., Savelsbergh, M., and Wang, X. (2012). Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research*, 223(2):295–303.

Agatz, N. A., Erera, A. L., Savelsbergh, M. W., and Wang, X. (2011). Dynamic ride-sharing: A simulation study in metro atlanta. *Transportation Research Part B: Methodological*, 45(9):1450–1464. Select Papers from the 19th ISTTT.

Alonso-González, M. J., van Oort, N., Cats, O., Hoogendoorn-Lanser, S., and Hoogendoorn, S. (2020). Value of

time and reliability for urban pooled on-demand services. *Transportation Research Part C: Emerging Technologies*, 115:102621.

Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., and Rus, D. (2017). On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences*, 114(3):462–467.

Balac, M., Hörl, S., and Axhausen, K. W. (2020). Fleet sizing for pooled (automated) vehicle fleets. *Transportation Research Record: Journal of the Transportation Research Board*, 2674(9):168–176.

Boesch, P. M., Ciari, F., and Axhausen, K. W. (2016). Autonomous vehicle fleet sizes required to serve different levels of demand. *Transportation Research Record: Journal of the Transportation Research Board*, 2542(1):111–119.

BV, T. I. (2024). Tomtom traffic index: Ranking 2024. `https://www.tomtom.com/traffic-index/ranking/?country=NL`. Accessed: 2024-07-15.

Bösch, P. M., Becker, F., Becker, H., and Axhausen, K. W. (2018). Cost-based analysis of autonomous mobility services. *Transport Policy*, 64:76–91.

Cats, O., Kucharski, R., Danda, S. R., and Yap, M. (2022). Beyond the dichotomy: How ride-hailing competes with and complements public transport. *PLOS ONE*, 17(1):1–17.

Engelhardt, R., Dandl, F., Syed, A.-A., Zhang, Y., Fehn, F., Wolf, F., and Bogenberger, K. (2022). Fleetpy: A modular open-source simulation tool for mobility on-demand services.

Fagnant, D. J. (2015). Shared autonomous vehicles: Model formulation, sub-problem definitions, implementation details, and anticipated impacts. *2015 American Control Conference (ACC)*, pages 2593–2593.

Fagnant, D. J. (2018). Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in austin, texas. *Transportation*, 45:143–158.

Fan, Q., van Essen, J. T., and Correia, G. H. (2023). 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.

Fielbaum, A. and Pudāne, B. (2024). Are shared automated vehicles good for public- or private-transport-oriented cities (or neither)? *Transportation Research Part D: Transport and Environment*, 136:104373.

Furuhata, M., Dessouky, M., Ordóñez, F., Brunet, M.-E., Wang, X., and Koenig, S. (2013). Ridesharing: The state-of-the-art and future directions. *Transportation Research Part B: Methodological*, 57:28–46.

Horni, A., Nagel, K., and Axhausen, K. W., editors (2016). *The Multi-Agent Transport Simulation MATSim*. Ubiquity Press, London. License: CC-BY 4.0.

Ishibashi, Y. and Akiyama, E. (2022). Predicting the impact of shared autonomous vehicles on tokyo transportation using matsim. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 3258–3265.

Jin, W.-L., Martinez, I., and Menendez, M. (2021). Compartmental model and fleet-size management for shared mobility systems with for-hire vehicles. *Transportation Research Part C: Emerging Technologies*, 129:103236.

KiM Netherlands Institute (2024). New values of travel time, reliability and comfort in the Netherlands.

Kucharski, R. and Cats, O. (2020). Exact matching of attractive shared rides (exmas) for system-wide strategic evaluations. *Transportation Research Part B: Methodological*, 139:285–310.

Kumakoshi, Y., Hanabusa, H., and Oguchi, T. (2021). Impacts of shared autonomous vehicles: Tradeoff between parking demand reduction and congestion increase. *Transportation Research Interdisciplinary Perspectives*, 12:100482.

Liu, A., Zhong, S., Sun, D., Gong, Y., Fan, M., and Song, Y. (2024). Joint optimal pricing strategy of shared autonomous vehicles and road congestion pricing: A regional accessibility perspective. *Cities*, 146:104742.

Militão, A. M. and Tirachini, A. (2021). Optimal fleet size for a shared demand-responsive transport system with human-driven vs automated vehicles: A total cost minimization approach. *Transportation Research Part A: Policy and Practice*, 151:52–80.

Monteiro, C. M., Machado, C. A. S., Lage, M. D. O., Berssaneti, F. T., Davis, C. A., and Quintanilha, J. A. (2021). Optimization of carsharing fleet size to maximize the number of clients served. *Computers, Environment and Urban Systems*, 87:101623.

Murtagh, E. M., Mair, J. L., Aguiar, E., Tudor-Locke, C., and Murphy, M. H. (2021). Outdoor walking speeds of apparently healthy adults: A systematic review and meta-analysis. *Sports Medicine*, 51(1):125–141.

Narayanan, S., Chaniotakis, E., and Antoniou, C. (2020). Shared autonomous vehicle services: A comprehensive review. *Transportation Research Part C: Emerging Technologies*, 111:255–293.

OpenStreetMap (2024). Openstreetmap [data set]. `https://www.openstreetmap.org`. Available as open data under the Open Data Commons Open Database License (ODbL).

ProRail (2025). Prorail: Homepage. Accessed: 2025-5-28.

Qu, B., Mao, L., Xu, Z., Feng, J., and Wang, X. (2022). How many vehicles do we need? fleet sizing for shared autonomous vehicles with ridesharing. *IEEE Transactions on Intelligent Transportation Systems*, 23(9):14594–14607.

Rodrigue, J.-P. (2024). *The Geography of Transport Systems*. Routledge, London, 6 edition. eBook, 402 pages.

Schröder, D. and Kaspi, M. (2024). Quantifying the external costs of autonomous on-demand ride pooling services. *Case Studies on Transport Policy*, 18:101302.

Seo, T. and Asakura, Y. (2022). Multi-objective linear optimization problem for strategic planning of shared autonomous vehicle operation and infrastructure design. *IEEE Transactions on Intelligent Transportation Systems*, 23(4):3816–3828.

Soza-Parra, J., Kucharski, R., and Cats, O. (2024). The shareability potential of ride-pooling under alternative spatial demand patterns. *Transportmetrica A Transport Science*, 20(2).

voor de Statistiek (CBS), C. B. and (RWS), R. (2015). Onderzoek Verplaatsingen in Nederland 2014 - OViN 2014.

Vosooghi, R., Puchinger, J., Jankovic, M., and Vouillon, A. (2019). Shared autonomous vehicle simulation and service design. *Transportation Research Part C: Emerging Technologies*, 107:15–33.

Wang, B., Ordonez Medina, S. A., and Fourie, P. (2018). Simulation of autonomous transit on demand for fleet size and deployment strategy optimization. *Procedia Computer Science*, 130:797–802.

Xia, J., Curtin, K. M., Li, W., and Zhao, Y. (2015). A new model for a carpool matching service. *PLOS ONE*, 10(6):1–23.

Zwick, F., Kuehnel, N., Moeckel, R., and Axhausen, K. W. (2021). Ride-pooling efficiency in large, medium-sized and small towns -simulation assessment in the munich metropolitan region. *Procedia Computer Science*, 184:662–667. The 12th International Conference on Ambient Systems, Networks and Technologies (ANT) / The 4th International Conference on Emerging Data and Industry 4.0 (EDI40) / Affiliated Workshops.

Zwick, F., Wilkes, G., Engelhardt, R., Axer, S., Dandl, F., Rewald, H., Kostorz, N., Fraedrich, E., Kagerbauer, M., and Axhausen, K. W. (2022). Mode choice and ride-pooling simulation: A comparison of mobitopp, fleetpy, and matsim. *Procedia Computer Science*, 201:608–613. The 13th International Conference on Ambient Systems, Networks and Technologies (ANT) / The 5th International Conference on Emerging Data and Industry 4.0 (EDI40).

# A Literature Review Supplementary Tables

Table 6: Comparison between the ExMAS algorithm and the time window-based method

| Comparison Dimension | ExMAS | Time Window-based |
|---|---|---|
| **Key Features** | Utility-based static demand optimization; provides a global, system-level perspective to support strategic decision-making. | Real-time matching of dynamic demand; focuses on short-term operational adjustments based on preset batch processing times. |
| **Advantages** | Supports systematic evaluation of the impact of different operational strategies by analyzing long-term demand patterns and optimizing the combined utility of passengers and operators. | Responds quickly to real-time demand changes, ensuring instant matching within the preset time window. |
| **Applicability for Strategy Evaluation** | The framework is suitable for evaluating system performance under different strategies (including cost, efficiency, user satisfaction, and the resulting resource requirements such as fleet size), assisting strategic decision-making, rather than being directly designed to output a single optimal fleet size. | Not designed for strategic evaluation; typically operates under a fixed fleet size to maximize short-term vehicle utilization. |
| **Matching Method** | Focuses on travel utility, dynamically balancing waiting time, detours, and system efficiency; suitable for static, reservation-based demand scenarios. | Matches based on strict time window constraints (e.g., 5 minutes), which may result in high levels of sharing with insufficient attractiveness. |
| **Application Scenarios** | Suitable for static demand analysis where all trips are known in advance, aiming to evaluate the impact of long-term strategies. | Most suitable for real-time operational scenarios requiring rapid response and instant matching. |

Table 7: Summary of Methods and Limitations

| Article | Method | Limitations |
|---|---|---|
| (Jin et al., 2021) | Mathematical modeling | Delays in matching and fleet size limit operation. Requires data calibration for critical densities in the fundamental diagram. |
| (Kucharski and Cats, 2020) | Mathematical modeling | The model assumes static demand and supply, ignores real-time traffic dynamics and demand elasticity, and does not consider fleet dispatching, rebalancing costs, or operational factors such as deadheading and driver incentives. |
| (Qu et al., 2022) | Mathematical modeling | Does not consider passenger satisfaction, vehicle charging and maintenance, supply-demand imbalance, repositioning of SAVs. |
| (Wang et al., 2018) | Simulation | The impact of different vehicle capacities was not considered, nor were passenger preferences for ride-sharing. |

Table 7: Summary of Methods and Limitations (continued)

| Article | Method | Limitations |
|---|---|---|
| (Seo and Asakura, 2022) | Mathematical modeling | The model cannot separately identify individual traveler path flows and routes; demand is unknown or uncertain, potentially overestimating SAVs efficiency. |
| (Monteiro et al., 2021) | Simulation + Mathematical modeling | Factors such as weather and traffic incidents, which could impact vehicle availability, are not considered in the simulation. |
| (Fan et al., 2023) | Mathematical modeling | Assumes that travel demand is known and fixed, does not incorporate passenger waiting time, considers only non-ride-pooling services, and simplifies the traffic environment. The model does not reflect the dynamic nature of real-world demand, ride-pooling mechanisms, or the complexity of multimodal transportation. |
| (Balac et al., 2020) | Simulation + Mathematical modeling | Considers only car trips within the city of Zurich as the source of demand, assumes static and a priori known demand, does not analyze mode shift or the formation of new demand, and does not consider external inflows or unpredictable travel events, which limits the applicability and realism of the conclusions. |
| (Militão and Tirachini, 2021) | Simulation | The study relies on simulation data from Munich and specific vehicle assignment strategies. The generalizability of the conclusions and their applicability to more complex scheduling algorithms require further validation. |
| (Vosooghi et al., 2019) | Simulation | Does not consider dynamic pricing strategies, the vehicle rebalancing algorithm requires further optimization, and the charging needs of electric SAVs and the layout of charging stations are not included in the simulation, which affects the practical applicability of the results. |
| (Boesch et al., 2016) | Simulation | Assumes static travel demand and idealized traffic conditions, does not consider ride-pooling, vehicle relocation, energy constraints, or demand hotspot optimization, and the conclusions are based only on the Zurich area, thus limiting regional applicability. |

## B Key Performance Indicators

Table 8: Fleet Size Metrics

| KPI Name | Description |
|---|---|
| Total Vehicles | Number of unique vehicles used to serve the rides |
| Peak Concurrent Vehicles | Maximum number of concurrently serving vehicles |

Table 9: Efficiency Metrics

| KPI Name | Description | Formula |
|---|---|---|
| Vehicle Reuse Rate | Percentage of vehicles utilized for multiple trips | $\|V_{reused}\|/\|V_{used}\|$<br>Where $V_{reused}$ is the set of vehicles serving more than one ride. |
| Average Rides per Vehicle | Average number of trips handled by each vehicle | $\|R^*\|/\|V_{used}\|$ |
| Saved Miles Ratio | Proportion of total vehicle mileage saved compared to the shortest path mileage for passengers, modification on pax detour ratio from(Jin et al., 2021)<br>(How much mileage is saved by ride-sharing compared to the shortest route) | $(D_{direct} - D_{service})/D_{direct}$<br>$D_{service} = \sum_{(r,v)\in R^*} d_{service}(r)$,<br>$D_{direct} = \sum_{p\in Pax_{served}} d_{direct}(p)$. |
| Extra Mileage Ratio | Proportion of pickup mileage to total mileage (Jin et al., 2021)<br>(Additional distance traveled for pickup / total distance) | $D_{pick}/(D_{pick} + D_{service})$<br>$D_{pick} = \sum_{(r,v)\in R^*} d_{pick}$. |
| Time Utilization Rate | Ratio of vehicle service time to total active time | $T_{service}/T_{active}$<br>$T_{service} = \sum_{(r,v)\in R^*} t_{service}(r)$,<br>$T_{active} = \sum_{v\in V_{used}} (t^{last}_{end}(v) - t^{first}_{start}(v))$. |
| Moving Time Ratio | Ratio of vehicle moving time (service time + pickup time) to total active time | $(T_{service} + T_{pick})/T_{active}$<br>$T_{pick} = \sum_{(r,v)\in R^*} t_{pick}$. |

Table 10: Pooling Metrics

| KPI Name | Description | Formula |
|---|---|---|
| Pooling Ratio | Percentage of trips involving ride-sharing in the selected solution | $\|Pax_{shared}\|/\|Pax_{served}\|$<br>$Pax_{shared}$ is the set of passengers who actually joined a shared ride, and $Pax_{served}$ is the set of all served passengers. |
| Shared Passengers Ratio | Proportion of passengers who was arranged to shared rides from the original rides optionsKucharski and Cats (2020) | $\|Pax_{shared\_options}\|/\|Pax_{all}\|$<br>$Pax_{shared\_options}$ is the set of passengers who were offered shared ride options. |

Table 11: Waiting Time Metrics

| KPI Name | Description | Formula |
|---|---|---|
| Average Vehicle Waiting Time | Average waiting time for vehicles to pick up passengers per ride | $\sum_{(r,v)\in R^*} t_{wait\_veh}/\lvert R^*\rvert$<br><br>$t_{wait\_veh} = \max(0, t_{actual\_start} - t_{arrival})$. |
| Average Passenger Waiting Time | Average waiting time for passengers before being picked up | $\sum_{p\in Pax_{served}} t_{wait}(p)/\lvert Pax_{served}\rvert$<br><br>$t_{wait}(p) = \max(0, t_{pickup\_actual}(p) - t_{req}(p))$. |
| Average Pickup Time | Average travel time for vehicles to pick up passengers | $\sum_{(r,v)\in R^*} t_{pick}/\lvert R^*\rvert$<br><br>$t_{pick}$ is the travel time for vehicle v from its current position to the origin of ride r. |

Table 12: KPIs Used in PT Analysis

| KPI Name | Description | Formula |
|---|---|---|
| Average Quit PT Distance | Average PT segment distance for passengers switched to PT | $\sum_{p\in Pax_{PT}} d_{PT}(p)/\lvert Pax_{PT}\rvert$ |
| Average Quit Distance | Average original requested travel distance for passengers switched to PT | $\sum_{p\in Pax_{PT}} d_{origin}(p)/\lvert Pax_{PT}\rvert$ |
| Walk Ratio | Proportion of passengers switched to direct walking | $\lvert Pax_{walk}\rvert/\lvert Pax_{all}\rvert$ |
| PT Ratio | Proportion of passengers switched to PT | $\lvert Pax_{PT}\rvert/\lvert Pax_{all}\rvert$ |

Table 13: Utility Metrics

| KPI Name | Description | Formula |
|---|---|---|
| Passenger Utility | Total disutility for served passengers (considering delays) | $\sum_{(r,v)\in R^*} U_{pax}(r,v)$<br><br>Where $U_{pax}(r,v)$ is the passenger disutility calculated when vehicle v serves ride r in the selected solution. |
| Vehicle Driving Cost | Cost associated with vehicle driving | $\sum_{(r,v)\in R^*} C_{drive}(r,v) \cdot f_{bal}$<br><br>$C_{drive}(r,v)$ is calculated based on the candidate solution, and $f_{bal}$ is the balance factor parameter. |
| Vehicle Waiting Cost | Total cost of vehicles waiting for passengers | $(\sum_{(r,v)\in R^*} t_{wait\_veh}) \cdot c_{wait\_veh} \cdot f_{bal}$<br><br>$t_{wait\_veh}$ is the waiting time after the vehicle arrives at the first passenger's origin, and $c_{wait\_veh}$ is the cost parameter per unit waiting time. |

## C Assumption Parameters

This study sets fixed values for parameters supported by research, such as value of time and price, and parameters that can be validated through the ODiN dataset. These parameters will not be included in the analysis, as they

serve as a basis for the system to accurately reflect reality(Bösch et al., 2018)(Alonso-González et al., 2020)(KiM Netherlands Institute, 2024)(Murtagh et al., 2021).

Table 14: Description of Fixed Input Parameters for the Model (SAV Operator and PT Parameters)

| **Parameter** | **Symbol** | **Unit** | **Default Value** | **Side** | **Description** | **Reference** |
|---|---|---|---|---|---|---|
| Average SAV Speed | $v_{avg}$ | m/s | 8 | SAV | Average speed of shared automated vehicles (SAVs). | Average speed from Amsterdam(BV, 2024), applied for all cities for comparability |
| Maximum Pickup Delay | $\Delta t^{p,\max}$ | s | 600 | SAV | Default value for the maximum pickup delay, aligning with real-world service levels. | Assumption |
| Shared Discount | $\delta$ | % | 30 | SAV | Default discount for shared rides, based on current platform strategies. | Assumption |
| SAV Price per km | $\pi$ | euros/km | 0.3714 | SAV | Per kilometer fare for solo SAV rides. | (Bösch et al., 2018) |
| PT Price per km | $\text{Fare}_{\text{dist}}$ | euros/km | 0.22 | PT | Per kilometer fare for public transport. | Estimation based on NS(Nederlandse Spoorwegen) fare system |
| PT Fixed Price | $\text{Fare}_{\text{fixed}}$ | euros | 1.12 | PT | Base fare for public transport. | Estimation based on NS fare system |
| Average PT Speed | $v_{PT}$ | m/s | 7 | PT | Average operating speed of public transport (rail). | Assumption |
| SAV Operational Cost | $c_t$ | euros/s | 0.0425 | SAV | Operator's cost per second for vehicle operation, covering both service and deadhead travel. | (Bösch et al., 2018) |
| Max degree (vehicle capacity) | – | – | 6 | SAV | Maximum passenger per ride. | Assumption |
| Balance factor | $f_{bal}$ | – | Calibrated | SAV | Balances operator costs and passenger utility. | Assumption |
| New Vehicle Fixed Cost | $c_f$ | euros | 20 | SAV | Fixed cost to activate a new vehicle. Estimated based on daily depreciation data for operator vehicles. | Assumption |
| Average Waiting Time for New Vehicles | $t_{wait}^{new}$ | s | 400 | SAV | Waiting time for a new vehicle to reach a passenger. | Assumption |

Table 15: Description of Fixed Input Parameters for the Model (User Parameters)

| Parameter Name | Unit | Default Value | Description | Reference |
|---|---|---|---|---|
| VoT commute in vehicle | euros/h | 10.8 | Value of time for commuters in vehicle. | (Alonso-González et al., 2020) |
| VoT commute waiting | euros/h | 12.06 | Value of time for commuters while waiting. | (Alonso-González et al., 2020) |
| VoT leisure waiting | euros/h | 9.25 | Value of time for non-commuters while waiting. | (Alonso-González et al., 2020) |
| VoT leisure in vehicle | euros/h | 9.94 | Value of time for non-commuters in vehicle. | (Alonso-González et al., 2020) |
| VoT walking | euros/h | 10.40 | Value of time for walking in multimodal trips. | (KiM Netherlands Institute, 2024) |
| VoT public transport | euros/h | 7.13 | Value of time for public transport in multimodal trips. | (KiM Netherlands Institute, 2024) |
| Average Walking Speed | m/s | 1.31 | Average walking speed for passengers. | (Murtagh et al., 2021) |
| Pax delay | s | 15 | Time for passengers to board or alight. | Assumption |
| $\lambda_{SAV}$ | - | 0.3 | The scale parameter for the SAV nest in the nested logit model, representing the degree of correlation between shared and non-shared SAV options. A value closer to 0 indicates higher correlation. | Assumption |

## D Urban Characteristic Indicators

This study establishes a comprehensive framework to analyze city characteristics through three key dimensions:

**Population and City Scale** Population ($P$) serves as the primary classifier for city scale:

$$\text{City Scale} = \begin{cases} \text{Large}, & P > 200,000 \\ \text{Medium}, & 100,000 \leq P \leq 200,000 \\ \text{Small}, & P < 100,000 \end{cases}$$

This classification provides the fundamental basis for comparing SAV deployment strategies across different urban contexts.

**Network Structure Indicators** These metrics characterize the urban road network topology, the road networks have been transformed into simple graph $G(V, E)$:

1. **Link Density** ($\rho_L$): Measures network coverage intensity

$$\rho_L = \frac{|E|}{A} \tag{9}$$

where $|E|$ is the number of road segments and $A$ is the urban area. Higher density indicates better network coverage and more routing options.

2. **Network Connectivity** ($C$): also known as the Gamma Index ($\gamma$). This metric quantifies the ratio of existing edges to the maximum possible edges in a planar graph, assessing the network's overall interconnectedness. It is calculated as:

$$C = \frac{|E|}{3|V| - 6} \tag{10}$$

A higher value of $C$, approaching 1, indicates a more densely connected network with greater accessibility and more routing options, which can be advantageous for efficient SAV deployment and operations. Conversely, values closer to 0 suggest a sparser network structure(Rodrigue, 2024).

3. **Average Clustering Coefficient ($\bar{C}$)**: A measure of network transitivity, this indicates the overall tendency for nodes to form local clusters. It is the average of local clustering coefficients ($C_i$), where $C_i$ quantifies how close a node's neighbors are to forming a clique.

$$C_i = \frac{2T_i}{d_i(d_i - 1)} \tag{11}$$

where $T_i$ is the number of triangles passing through node $i$, and $d_i$ is the degree of node $i$. If $d_i < 2$, $C_i$ is defined as 0.

The average clustering coefficient is then:

$$\bar{C} = \frac{1}{|V|} \sum_{i \in V} C_i \tag{12}$$

A higher average clustering coefficient (closer to 1) indicates that, on average, the neighbors of nodes are densely interconnected, signifying a strong presence of local clustered structures within the network(Rodrigue, 2024).

4. **Network Meshedness** ($M$): This index, also known as the Alpha Index ($\alpha$) for planar networks, quantifies the degree of circuit redundancy or cyclicity within the network. It is calculated as the ratio of the actual number of fundamental circuits (or cycles) to the maximum possible number of circuits in a planar graph with the same number of nodes(Rodrigue, 2024). It is calculated as:

$$M = \frac{|E| - |V| + 1}{2|V| - 5} \tag{13}$$

A higher $M$ indicates a more grid-like topology with greater routing flexibility and redundancy, potentially benefiting SAV dispatching and resilience.

5. **Average Node Degree** ($\bar{d}$): This metric represents the average number of edges incident to a node (intersection) in the network. It provides a measure of the local connectivity at intersections.It is calculated as:

$$\bar{d} = \frac{1}{|V|} \sum_{i \in V} d_i \tag{14}$$

A higher $\bar{d}$ implies more immediate directional choices at intersections, offering increased local routing options. However, overall network navigability also depends on global topological properties.

**Demand Pattern Indicators** These metrics reveal travel demand characteristics:

1. **Commuting Trip Ratio** ($r_c$): Measures proportion of commuting trips

$$r_c = \frac{|Q_c|}{|Q|} \tag{15}$$

where $Q_c$ represents commuting trips and $Q$ total trips. Higher ratio indicates stronger commuting patterns.

2. **Average Commuting Distance** ($\bar{d_c}$): Reflects urban sprawl

$$\bar{d_c} = \frac{1}{|Q_c|} \sum_{i \in Q_c} d_i \tag{16}$$

Longer distances suggest greater potential for ride-sharing benefits.

# Appendix B Nomenclature

Table B.1: Nomenclature of Variables and Parameters

| Symbol | Description |
|---|---|
| $\alpha$ | A factor to balance operator costs and passenger utility in the joint cost function. |
| $\beta_{ivt}$ | Value of time in vehicle for SAVs. |
| $\beta_{PT,ivt}$ | Value of time for in-vehicle travel on public transport. |
| $\beta_{wait}$ | Value of time for waiting for SAVs. |
| $\beta_{walk}$ | Value of time for walking. |
| $c_f$ | The fixed cost incurred when activating a new vehicle. |
| $c_t$ | Operator's cost per second for vehicle operation, covering both service and deadhead travel. |
| $c_w$ | The unit time cost of operator incurred by a vehicle waiting for passengers due to early arrival. |
| $c_{r,v}$ | The total joint cost of a ride-vehicle pair $(r, v)$. |
| $d_{PT}$ | The main-line travel distance on the public transport network. |
| $\delta$ | The fare discount percentage for passengers when choosing shared rides. |
| $D$ | Total travel demand. |
| $\delta_{PT}$ | The discount percentage applied to the total fare of public transport. |
| $\Delta t^{ba}$ | The fixed time allocated for each passenger's boarding and alighting. |
| $\Delta t^{p}_{i,r}$ | The pickup delay experienced by passenger $i$ in shared ride $r$. |
| $\Delta t^{p,\max}$ | A critical service level parameter defining the maximum pickup delay a passenger can tolerate. |
| $\Delta t^{wait}_{i}$ | The initial waiting time for passenger $i$ before the ride begins. |
| $f_{bal}$ | An internal factor to balance the numerical scales of operator cost and user disutility. |
| $f$ | Average driving frequency per person per day. |
| $F^{ns}_{i}$ | The fare for a non-shared ride for passenger $i$. |
| $F^{s}_{i,r}$ | The fare for a shared ride $r$ for passenger $i$. |
| $\text{Fare}_{\text{dist}}$ | The distance-based component of the public transport fare. |
| $\text{Fare}_{\text{fixed}}$ | The fixed component of the public transport fare. |
| $\mathbb{I}(\text{new}_v)$ | An indicator function that is 1 if vehicle $v$ is newly activated, and 0 otherwise. |
| $l_i$ | The travel distance for trip $i$ in meters. |
| $\lambda_{SAV}$ | Scale parameter for the SAV nest in the nested logit model. |
| $\omega$ | A factor reflecting passengers' resistance to the additional time costs associated with shared rides. |
| $O_{r,v}$ | The total operator cost for assigning vehicle $v$ to serve ride $r$. |
| $p$ | An individual passenger travel request. |
| $P$ | The complete set of all passenger requests to be served. |
| Pop | The population of a city. |
| $\pi$ | The per-kilometer fare for solo SAV rides. |
| Pairs | The set of all feasible (ride, vehicle) assignment pairs. |
| $\text{Pairs}_p$ | The subset of candidate pairs that include passenger $p$. |
| $r$ | A specific ride, composed of a unique set of passengers. |
| $s$ | The percentage of simulated demand relative to the total actual travel volume. |
| $t_i$ | The direct, non-shared in-vehicle travel time for passenger $i$. |
| $\hat{t}_{i,r}$ | The in-vehicle travel time for passenger $i$ in shared ride $r$. |
| $t_{PT}$ | The in-vehicle travel time on public transport. |
| $t_{r,v}$ | The total driving time for vehicle $v$ to complete ride $r$. |
| $t^{w}_{r,v}$ | The waiting time of vehicle $v$ before starting ride $r$ due to early arrival. |
| $t^{new}_{wait}$ | The preset average waiting time for a newly activated vehicle to arrive. |
| $t_{\text{walk}}$ | The total walking time to and from public transport stations. |
| $U_r$ | The total user disutility for all passengers in ride $r$. |
| $U_{PT}$ | Total generalized cost (disutility) of a trip using public transport. |
| $U^{ns}_{i}$ | The total disutility for passenger $i$ in a non-shared ride. |
| $U^{s}_{i,r}$ | The total disutility for passenger $i$ in a shared ride $r$. |
| $v$ | An individual vehicle from the available fleet. |
| $v_{avg}$ | The average speed of SAVs, used to calculate vehicle travel times and various time-related KPIs. |
| $v_{PT}$ | The average operational speed of public transport services. |
| $x_{r,v}$ | Binary variable; 1 if the ride-vehicle pair is selected, 0 otherwise. |
| $VoT$ | Generalized value of time, simplified for discussion. |
| $WtSR$ | Simplification for Willingness-to-Share Resistance Factor, same as $\omega$. |

# Appendix C Fixed Parameters

In addition to the variable parameters analyzed in the experimental scenarios, several core parameters of the model were held constant to maintain a focused and comparable analysis. The values for these parameters, along with their justifications, are summarized in Table C.1.

Table C.1: Description of Fixed Input Parameters for the Model (SAV Operator and PT Parameters)

| **Parameter** | **Symbol** | **Unit** | **Default Value** | **Side** | **Description** | **Reference** |
|---|---|---|---|---|---|---|
| Average SAV Speed | $v_{avg}$ | m/s | 8 | SAV | Average speed of shared automated vehicles (SAVs). | Average speed from Amsterdam(BV, 2024), applied for all cities for comparability |
| Maximum Pickup Delay | $\Delta t^{p,\max}$ | s | 600 | SAV | Default value for the maximum pickup delay, aligning with real-world service levels. | Assumption |
| Shared Discount | $\delta$ | % | 30 | SAV | Default discount for shared rides, based on current platform strategies. | Assumption |
| SAV Price per km | $\pi$ | euros/km | 0.3714 | SAV | Per kilometer fare for solo SAV rides. | (Bösch et al., 2018) |
| PT Price per km | $\text{Fare}_{\text{dist}}$ | euros/km | 0.22 | PT | Per kilometer fare for public transport. | Estimation based on NS(Nederlandse Spoorwegen) fare system |
| PT Fixed Price | $\text{Fare}_{\text{fixed}}$ | euros | 1.12 | PT | Base fare for public transport. | Estimation based on NS fare system |
| Average PT Speed | $v_{PT}$ | m/s | 7 | PT | Average operating speed of public transport (rail). | Assumption |
| SAV Operational Cost | $c_t$ | euros/s | 0.0425 | SAV | Operator's cost per second for vehicle operation, covering both service and deadhead travel. | (Bösch et al., 2018) |
| Max degree (vehicle capacity) | – | – | 6 | SAV | Maximum passenger per ride. | Section D |
| Balance factor | $f_{bal}$ | – | Calibrated | SAV | Balances operator costs and passenger utility. | Assumption |
| New Vehicle Fixed Cost | $c_f$ | euros | 20 | SAV | Fixed cost to activate a new vehicle. Estimated based on daily depreciation data for operator vehicles. | Assumption |

Continued on next page

**Table C.1 – continued from previous page**

| Parameter | Symbol | Unit | Default Value | Side | Description | Reference |
|---|---|---|---|---|---|---|
| Average Waiting Time for New Vehicles | $t_{wait}^{new}$ | s | 400 | SAV | Waiting time for a new vehicle to reach a passenger. | Assumption |

Table C.2: Description of Fixed Input Parameters for the Model (User Parameters)

| Parameter Name | Unit | Default Value | Description | Reference |
|---|---|---|---|---|
| VoT commute in vehicle | euros/h | 10.8 | Value of time for commuters in vehicle. | (Alonso-González et al., 2020) |
| VoT commute waiting | euros/h | 12.06 | Value of time for commuters while waiting. | (Alonso-González et al., 2020) |
| VoT leisure waiting | euros/h | 9.25 | Value of time for non-commuters while waiting. | (Alonso-González et al., 2020) |
| VoT leisure in vehicle | euros/h | 9.94 | Value of time for non-commuters in vehicle. | (Alonso-González et al., 2020) |
| VoT walking | euros/h | 10.40 | Value of time for walking in multimodal trips. | (KiM Netherlands Institute, 2024) |
| VoT public transport | euros/h | 7.13 | Value of time for public transport in multimodal trips. | (KiM Netherlands Institute, 2024) |
| Average Walking Speed | m/s | 1.31 | Average walking speed for passengers. | (Murtagh et al., 2021) |
| Pax delay | s | 15 | Time for passengers to board or alight. | Assumption |
| $\lambda_{SAV}$ | - | 0.3 | The scale parameter for the SAV nest in the nested logit model, representing the degree of correlation between shared and non-shared SAV options. A value closer to 0 indicates higher correlation. | Assumption |

# Appendix D Justification of Maximum Vehicle Capacity

In this study, the vehicle capacity is fixed at six passengers. This decision is grounded in the specific hierarchical mechanism used for generating ride-pooling candidates within the simulation framework. The system constructs ride-pooling options iteratively: it first generates all feasible two-person shareable rides, and then uses these as a basis to build three-person rides, and so on, up to the defined capacity limit.

A key implication of this hierarchical approach is that the generation of lower-occupancy rides is independent of the upper capacity limit. Therefore, altering the maximum capacity only truncates the generation process at a different level, without affecting the set of feasible ride candidates generated below that new limit.

To empirically validate this and quantify the impact of the maximum ride size on system-level performance, a robustness analysis was conducted. Using the dataset for a representative large city(Amsterdam) and keeping all other parameters at their default values, the simulation was run with the maximum number of passengers per ride (*max degree*) set from 2 up to 6. This allows for a clear distinction between the potential ride candidates generated by the algorithm and the final rides selected by the global cost optimization. The key results, including fleet efficiency and service level indicators, are summarized in Table D.1.

Table D.1: Robustness Analysis of Maximum Capacity in Amsterdam

| **Max. Capacity** | **2** | **3** | **4** | **5** | **6 (Default)** |
|---|---|---|---|---|---|
| **Potential Ride Candidates** | | | | | |
| Rides with 1 passenger | 6,705 | 6,705 | 6,705 | 6,705 | 6,705 |
| Rides with 2 passengers | 2,733 | 2,733 | 2,733 | 2,733 | 2,733 |
| Rides with 3 passengers | 0 | 11 | 11 | 11 | 11 |
| Rides with 4+ passengers | 0 | 0 | 0 | 0 | 0 |
| **Selected Rides (Final Solution)** | | | | | |
| Rides with 1 passenger | 4,941 | 4,968 | 4,968 | 4,968 | 4,968 |
| Rides with 2 passengers | 882 | 867 | 867 | 867 | 867 |
| Rides with 3 passengers | 0 | 1 | 1 | 1 | 1 |
| Rides with 4+ passengers | 0 | 0 | 0 | 0 | 0 |
| **Key System KPIs** | | | | | |
| Total Vehicles Used | 337 | 338 | 338 | 338 | 338 |
| Avg. Pax. Waiting Time (s) | 57.5 | 57.8 | 57.8 | 57.8 | 57.8 |
| Vehicle Reuse Rate (%) | 95.5% | 97.0% | 97.0% | 97.0% | 97.0% |

The robustness analysis demonstrates that key system performance indicators converge and remain stable for all capacity settings of three and above. The minor discrepancy at a capacity of two is an artifact of the framework's myopic and sequential vehicle assignment process, as detailed in Section 6.1.1, where the introduction of higher-order candidates to the initial pool alters the set of vehicle-ride pairings passed to the final optimizer. This finding consistently holds across all experimental scenarios in this study, confirming that a capacity of six is a non-restrictive choice that allows the analytical focus to remain on more influential strategic parameters.

# Appendix E Supplementary Key Performance Indicators

Table E.1: Definitions and Formulas of Supplementary KPIs

| KPI Name | Description | Formula |
|---|---|---|
| **Supplementary Fleet Size and Utility Metrics** | | |
| Vehicles Difference | The difference between initial fleet size and final used vehicles | $N_{init} - V_{used}$<br><br>$N_{init} = \lfloor \lvert Pax_{req} \rvert / p_{init} \rfloor$, where $p_{init}$ is the initial proportion parameter. |
| Passenger Utility | Total disutility for served passengers (considering delays) | $\sum_{(r,v)\in R^*} U_{pax}(r,v)$<br><br>Where $U_{pax}(r,v)$ is the passenger disutility calculated when vehicle v serves ride r in the selected solution. |
| Vehicle Driving Cost | Cost associated with vehicle driving | $\sum_{(r,v)\in R^*} C_{drive}(r,v) \cdot f_{bal}$<br><br>$C_{drive}(r,v)$ is calculated based on the candidate solution, and $f_{bal}$ is the balance factor parameter. |
| Vehicle Waiting Cost | Total cost of vehicles waiting for passengers | $(\sum_{(r,v)\in R^*} t_{wait\_veh}) \cdot c_{wait\_veh} \cdot f_{bal}$<br><br>$t_{wait\_veh}$ is the waiting time after the vehicle arrives at the first passenger's origin, and $c_{wait\_veh}$ is the cost parameter per unit waiting time. |

# Appendix F PT Network Data

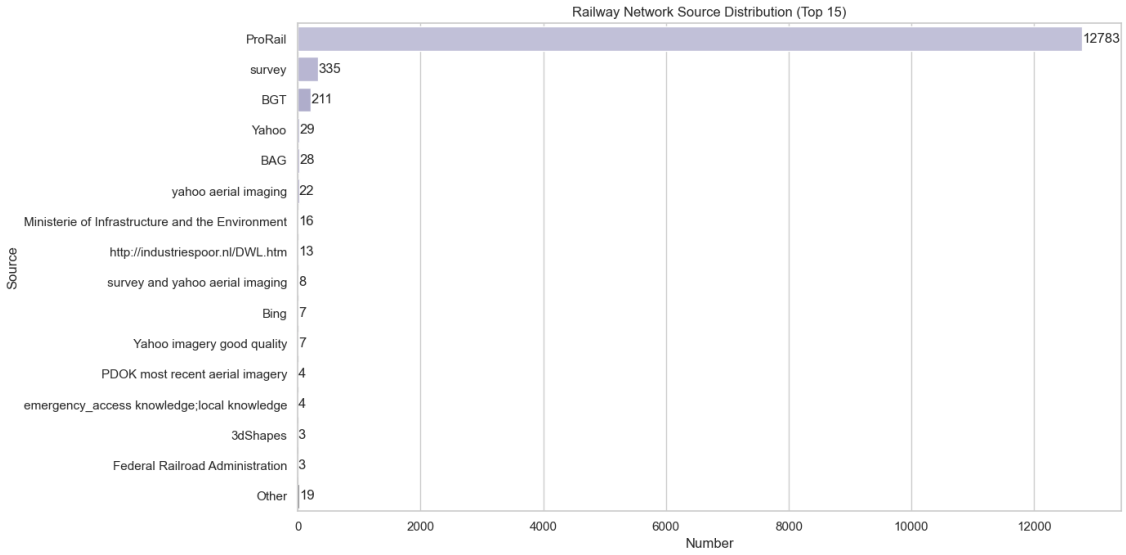
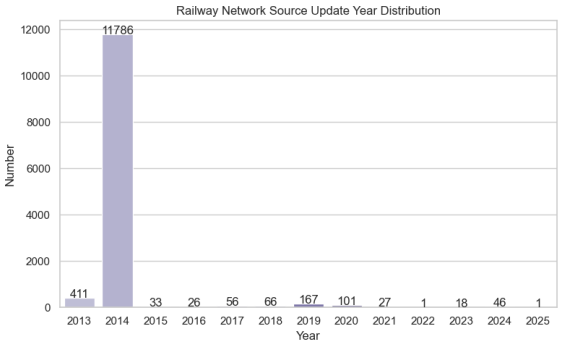
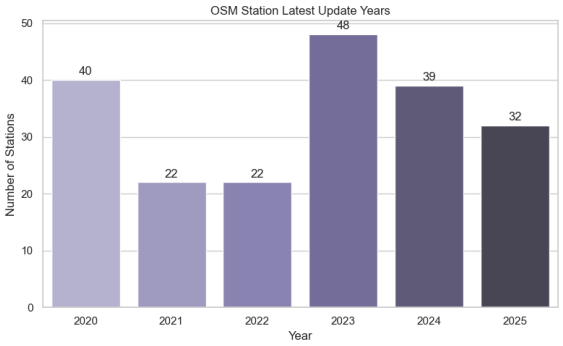



Figure F.1: Distribution of Main Data Sources for OSM Dutch Railway Network




(a) Update Year Distribution of OSM Dutch Railway Network Data




(b) Update Year Distribution of OSM Dutch Railway Station Data

Figure F.2: Update year distributions for OSM Dutch railway network and station data.

Almost all railway track data in the Netherlands within OSM were uniformly imported by ProRail, the Dutch infrastructure management companyProRail (2025). As shown in Figure F.2a, all railway network data used in this study were sourced from OSM after 2013, meaning that the track geometry and connectivity reflect the network status as of 2013 or later. In addition, metadata statistics indicate that all railway station data were updated after 2020 (Figure F.2b), ensuring that the locations and attributes of stations are current. Therefore, the analysis is based on a network topology no older than 2013, with station information reflecting updates through 2020, which together ensure the timeliness and reliability of the data.

# Appendix G Demand Generation

## G.1 Region Division and Node Classification

To accurately simulate commuting behavior (for instance, morning peak travel from residential to work areas and vice versa during the evening peak), it is essential to first identify functional zones within the urban space. This study leverages the geospatial data provided by OSM, utilizing the *osmnx*Boeing (2025) library to extract polygon information related to land use and points of interest, which serves as the basis for zone classification.

The classification approach focuses on identifying main urban functional units, as outlined below:

- **Residential Areas:** This group includes areas mainly used for housing. These are identified using OSM tags for residential land use (like *landuse=residential*) and groups of residential buildings (e.g., *building=apartments*, *building=house*).
- **Work Areas:** This group covers various workplaces. It includes office areas (tagged as *building=office*, *office=true*), industrial zones (*landuse=industrial*), as well as educational institutions and hospitals (tagged as *landuse=university*, *building=hospital*).
- **Other Areas:** Areas that do not clearly fit into residential or work categories—such as parks, empty lots, retail shops, restaurants, leisure spots, or zones with mixed functions—are grouped as other areas.

After classifying the zone polygons, each node in the road network from OSM is given a functional label (residential, work, or other) based on the zone type where its geographic coordinates fall. This node classification based on location helps in creating realistic commuting and non-commuting travel demand in later steps.

To show the results of zone division and node classification clearly, Figure G.1 uses Amsterdam as an example, displaying the identified functional zones and the related road network node distribution.

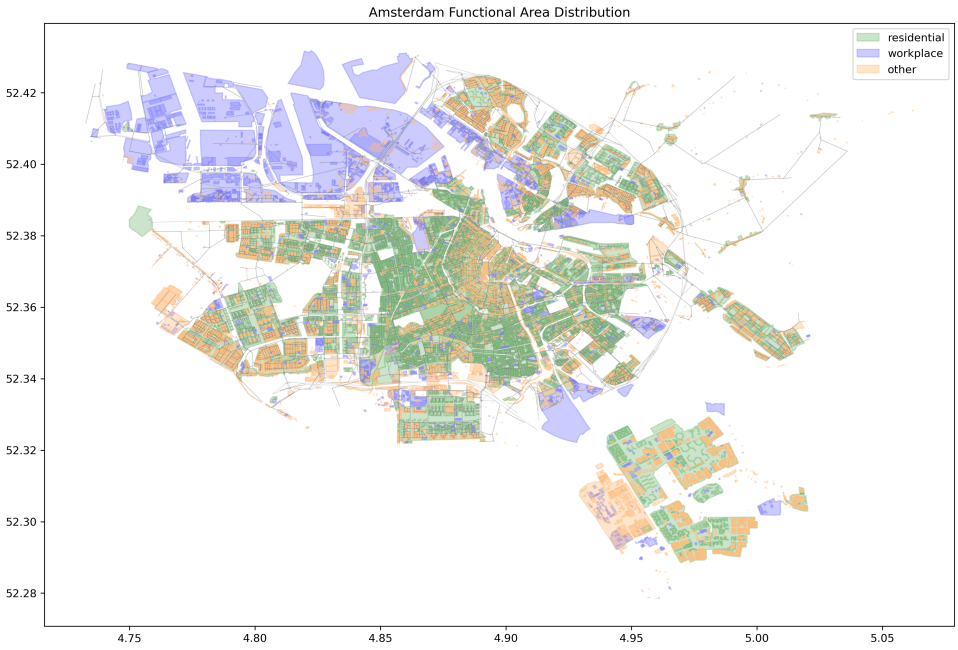



Figure G.1: Example of Functional Zone Division and Road Network Node Classification in Amsterdam

## G.2 Calibration of Travel Spatio-Temporal Characteristics

To reflect the real demand characteristics of cities, this study directly utilizes the travel data provided by the ODiN dataset to calibrate travel characteristics. These characteristics include the proportion of hourly trips to the total daily trips, the proportion of commuting trips within each hour's total trips,

the distribution of all trip distances, and the distribution of commuting trip distances. The following statistical results are obtained after calibration, using Amsterdam as a representative case.

The calculation of the four key statistical distributions is as follows:

- **Hourly Trip Proportion ($P_h$):** The proportion of total daily trips starting in hour $h$ is calculated as:
$$P_h = \frac{N_h}{N} \tag{G.1}$$
where $N_h$ is the number of trips starting in hour $h$, and $N$ is the total number of trips in the city's filtered dataset,

- **Hourly Commuting Trip Proportion ($S_h$):** The proportion of trips in hour $h$ that are for commuting purposes is given by:
$$S_h = \frac{N'_h}{N_h} \tag{G.2}$$
where $N'_h$ denotes the number of trips for commuting purposes starting in hour $h$, and $N_h$ is the total number of trips starting in the same hour.

- **Trip Distance Distribution ($P_c$):** The proportion of trips falling into a specific distance category $c$, is calculated as:
$$P_c = \frac{N_c}{N} \tag{G.3}$$
where $N_c$ is the number of trips in distance category $c$, the categories $c \in \{1, ..., 8\}$ correspond to the following ranges: 1 (0–0.5 km), 2 (0.5–1 km), 3 (1–2.5 km), 4 (2.5–3.7 km), 5 (3.7–5 km), 6 (5–7.5 km), 7 (7.5–10 km), and 8 (10–15 km).

- **Commuting Trip Distance Distribution ($S_c$):** Similarly, the distance distribution for the subset of commuting trips is:
$$S_c = \frac{N'_c}{N'_{\text{dist}}} \tag{G.4}$$
where $N'_c$ is the number of commuting trips in distance category $c$, and $N'_{\text{dist}}$ is the total number of commuting trips within the considered distance range.

Taking Amsterdam as an example, Figure G.2 illustrates the hourly number of total and commuting trips, as extracted from the ODiN dataset. The figure clearly reveals a typical urban bimodal commuting pattern: during the morning peak (6:00-9:00) and evening peak (16:00-19:00), commuting trips account for a significant proportion of the total trips. In contrast, during off-peak hours and nighttime, the proportion of commuting trips is relatively lower. By replicating such spatio-temporal distribution patterns, it can be ensured that the simulated demand reflects the tidal traffic flow patterns of daily urban operations. This is crucial for subsequently evaluating the operational efficiency of the SAV system across different time periods.

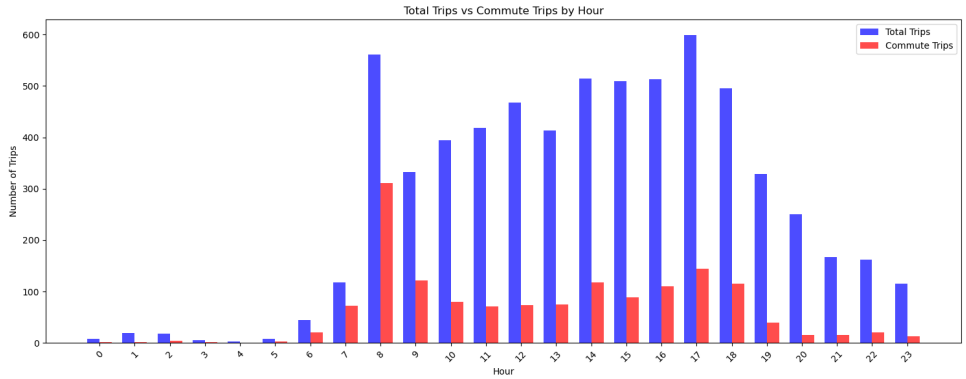



Figure G.2: Proportion of Commuting Trips in Amsterdam

In addition to the temporal distribution, the spatial characteristics of trip distances are also a critical aspect of calibration. Figure G.3 compares the distance distribution of the final simulated travel demand

in Amsterdam with the original distance distribution extracted from the ODiN dataset. The figure shows that the proportion of simulated distance distribution across various intervals largely aligns with the original data. In the short to medium distance ranges (e.g., 1-5 kilometers), the consistency between the generated demand and the original data is notably high, the short distance trips are dominant, reflecting the characteristics of intra-urban travel. However, in longer distance ranges, particularly 5-7.5 kilometers, the proportion of simulated demand is lower than that in the original dataset. This discrepancy arises because the origin of each trip is randomly selected without considering the topological structure of the road network, for example, the other type trips could chose an origin node in the corner of road network, with limited accessiblity to other nodes. Consequently, a destination corresponding to the specific distance may not exist on the network, leading to deviations in trip distances. Despite these differences, overall, the simulated demand distance distribution effectively captures the primary spatial scale characteristics of urban travel, providing a representative foundation for subsequent analysis.

Taking Amsterdam as an example, Figure G.3 presents a comparison between the actual simulated trip distance distribution and the distance distribution from the original dataset.

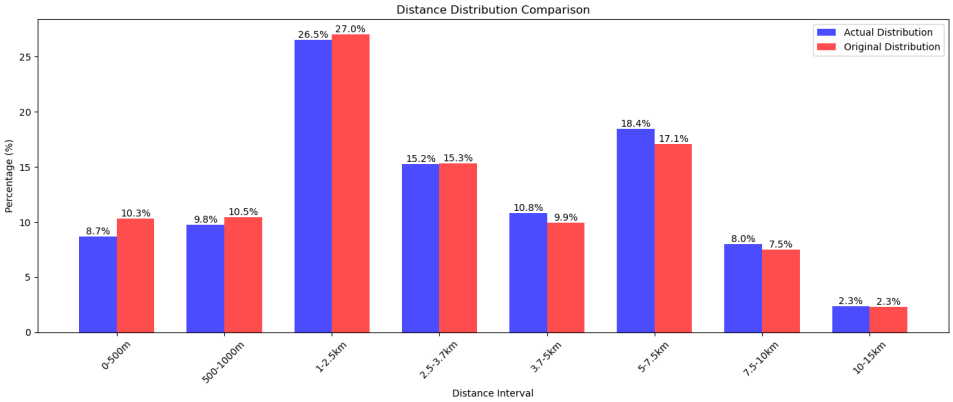



Figure G.3: Comparison of Trip Distance Distribution in Amsterdam with Real Data

Through the calibration of temporal and spatial characteristics, a solid demand data foundation for the following simulation of the SAV system has been generated, ensuring that the analysis results better reflect the potential performance of SAVs in specific urban environments.

# Appendix H City Characteristics and WtSR Calibration

Table H.1: City Characteristics and WtS/Ratio Comparison

| City Scale | City | Best WtS | Population | Simulated Shared Ratio | Real Shared Ratio |
|---|---|---|---|---|---|
| large | Amsterdam | 1.06 | 931298 | 0.55 | 0.56 |
| large | Rotterdam | 1.04 | 670610 | 0.54 | 0.53 |
| large | Den Haag | 1.06 | 566221 | 0.56 | 0.56 |
| large | Utrecht | 1.02 | 374238 | 0.52 | 0.51 |
| large | Eindhoven | 1.03 | 246417 | 0.52 | 0.51 |
| large | Groningen | 0.99 | 243768 | 0.52 | 0.53 |
| large | Tilburg | 1.04 | 229836 | 0.50 | 0.48 |
| large | Almere | 1.00 | 226500 | 0.53 | 0.54 |
| medium | Nijmegen | 1.02 | 187049 | 0.45 | 0.46 |
| medium | Apeldoorn | 0.98 | 168211 | 0.51 | 0.52 |
| medium | Haarlem | 0.99 | 167636 | 0.58 | 0.57 |
| medium | Arnhem | 1.01 | 167632 | 0.54 | 0.54 |
| medium | Haarlemmermeer | 0.94 | 163128 | 0.50 | 0.51 |
| medium | Amersfoort | 0.98 | 161852 | 0.56 | 0.56 |
| medium | Enschede | 0.98 | 161738 | 0.55 | 0.54 |
| medium | Zaanstad | 0.98 | 161389 | 0.52 | 0.51 |
| medium | Zwolle | 0.96 | 133141 | 0.61 | 0.61 |
| medium | Leiden | 0.90 | 130108 | 0.58 | 0.58 |
| medium | Leeuwarden | 0.98 | 128810 | 0.51 | 0.51 |
| medium | Zoetermeer | 0.95 | 128434 | 0.57 | 0.58 |
| medium | Maastricht | 0.98 | 125285 | 0.53 | 0.53 |
| medium | Ede | 0.96 | 123532 | 0.50 | 0.50 |
| medium | Westland | 0.97 | 115941 | 0.51 | 0.51 |
| medium | Alkmaar | 0.90 | 112304 | 0.55 | 0.55 |
| medium | Delft | 0.95 | 109577 | 0.50 | 0.52 |
| medium | Venlo | 0.99 | 103789 | 0.48 | 0.49 |
| medium | Deventer | 0.94 | 103405 | 0.53 | 0.54 |
| small | Purmerend | 0.95 | 95168 | 0.50 | 0.51 |
| small | Amstelveen | 0.98 | 95014 | 0.52 | 0.52 |
| small | Lelystad | 0.92 | 84080 | 0.55 | 0.55 |
| small | Veenendaal | 0.86 | 69440 | 0.55 | 0.53 |
| small | Zeist | 0.96 | 66641 | 0.49 | 0.49 |
| small | Nieuwegein | 0.98 | 65971 | 0.48 | 0.48 |
| small | Lansingerland | 0.98 | 65594 | 0.50 | 0.50 |
| small | Woerden | 0.90 | 53724 | 0.49 | 0.49 |
| small | Houten | 0.89 | 50847 | 0.51 | 0.51 |
| small | Noordoostpolder | 0.80 | 50035 | 0.53 | 0.53 |

It is worth noting that $\omega$ value calibrated against ODiN data in Section 5.1.1 aims to quantify the passenger's resistance to the total time lost associated with sharing rides (including extra travel time and potential departure delays) relative to their value of time, expressed as a weight factor. A higher $\omega$ indicates a lower tolerance for time inconvenience in exchange for price discounts; a lower $\omega$ implies a higher tolerance for such inconvenience. However, the calibration based on actual sharing rates reveals that large cities, where actual sharing rates are generally high, require higher $\omega$ values to match the ODiN data, whereas small cities require lower $\omega$ values.

In order to investigate the underlying reasons for the heterogeneity of passenger $\omega$ across different city backgrounds, it is necessary to examine other relevant parameters in the model, particularly the value of time parameter ($VoT$), which may potentially influence the calibration results of $\omega$. In this study, $VoT$ represents the economic measure of travel time costs, while $\omega$ focuses on quantifying the psychological

perception costs arising from the additional time incurred by choosing ride-pooling. Within the model, $\omega$ and $VoT$ jointly determine the acceptance of passengers towards ride-pooling services and the associated inconveniences.

Regarding parameter settings, this study specifies four types of waiting times for commuting, non-commuting, waiting, and in-vehicle scenarios. However, it should be noted that all cities under study adopt a unified $VoT$ value. Therefore, the observed differences in $\omega$ across cities during model calibration may partly result from insufficient consideration of the actual inherent differences in $VoT$ among these cities. This premise leads to the core question of this section: to what extent are the effects captured by the calibrated $\omega$ parameter independent of $VoT$? Is this parameter merely a compensatory adjustment arising from the lack of refined modeling of $VoT$ (for example, neglecting the real heterogeneity in $VoT$ caused by economic level or traffic conditions), or does it indeed capture and represent a purely subjective preference for the inconvenience brought by ride-pooling, independent of the general value of time? Furthermore, under the current $VoT$ setting, which may not fully capture the real heterogeneity across cities, it remains unknown whether the absolute value of $\omega$ obtained through model calibration can directly and accurately reflect the pure reference value defined in theory, that is, the true subjective assessment of inconvenience by passengers. Therefore, this section proceeds with a multiple regression analysis, as well as a single-parameter regression analysis for $VoT$, to explore the relationship between $\omega$ and $VoT$.

Table H.2: Comparison of regression results for ride-pooling rate (VoT only vs. VoT+$\omega$)

| Model | $N$ | $R^2$ | Adjusted $R^2$ | AIC | BIC | Parameter | Coefficient ($p$-value) |
|---|---|---|---|---|---|---|---|
| VoT only | 21 | 0.971 | 0.970 | -94.91 | -92.82 | VoT change | -221.18 (<0.001) |
| | | | | | | Constant | 0.54 (<0.001) |
| VoT+$\omega$ | 110 | 0.992 | 0.992 | -716.8 | -708.7 | VoT change | -227.59 (<0.001) |
| | | | | | | $\omega$ | -2.60 (<0.001) |
| | | | | | | Constant | 2.99 (<0.001) |

Table H.2 shows that, compared to the effect of $VoT$ alone on the ride-pooling rate ($R^2 = 0.971$), the combined effect of $\omega$ and $VoT$ achieves a better fit ($R^2 = 0.992$), and the improvement is significant. This provides some evidence for the rationality of $\omega$ as an independent parameter, indicating that it may capture behavioral dimensions related to ride-pooling that are not fully covered by $VoT$ in the model. It should be noted that, although the multiple regression analysis demonstrates the independent explanatory power of $\omega$, the complete independence of $\omega$ and $VoT$ at the behavioral level cannot be strictly confirmed due to the limitations of the model structure and parameter settings. Therefore, the absolute value of $\omega$ should be interpreted as a comprehensive reflection of subjective preferences within the current modeling framework, rather than the sole mapping of its theoretical definition.

Overall, the role of $\omega$ in the current model framework can be understood as dual. On the one hand, under the condition that the $VoT$ parameter is simplified (for example, adopting a unified value for all cities without reflecting their inherent heterogeneity), the calibrated value of $\omega$ may indeed absorb and compensate for the real differences in $VoT$ among different cities or groups that are not explicitly modeled, which can be regarded as a balancing mechanism in the model calibration process. On the other hand, even after considering this potential compensatory effect, $\omega$ parameter still has the potential to capture the subjective assessment of psychological and cultural factors associated with ride-pooling.

In ideal modeling practice, $VoT$ should be independently and finely calibrated based on more direct and granular observational data to reflect the real value of time for different cities and groups. This study indirectly observes and calibrates $\omega$ parameter through the ride-pooling rate. Due to the lack of other data sources, it is difficult to verify the true value of $\omega$ corresponding to its definition using data after precise $VoT$ calibration. Future research can use SP or travel behavior choice data to first calibrate VoT, and then calibrate $\omega$ separately through the ride-pooling rate, so that the pattern of $\omega$ across cities will have greater statistical significance. Therefore, the finding in this study that $\omega$ is higher in large cities and lower in small cities at the calibration stage should not be attributed to a lower willingness to share in large cities, but may also result from the generally higher $VoT$ among residents in large cities(Melo and Graham, 2018), and $\omega$ in fact captures the effect of the lack of refined modeling of $VoT$ in the model.

This study thus involves a common modeling issue, namely, when a model contains multiple parameters with similar or overlapping scopes, all aiming to characterize and quantify the subjective preferences of the research object, how can it be ensured during the calibration of a specific parameter that its value

does not capture effects that should be reflected by other related parameters? If, due to the limitations of the model structure, insufficient data, or calibration techniques, it is not possible to achieve absolute and clear separation of the effects among these subjective preference parameters in the model, then to what extent can the absolute value of any single parameter obtained through calibration still be regarded as a perfect and unbiased mapping of its pure theoretical definition? The answer often requires critical consideration. In such cases, the absolute value of the parameter itself may not be a direct and transparent reflection of its theoretical definition, and its interpretation must be closely integrated with the overall structure of the model, the specific calibration process, the characteristics of the data used, and the complex potential associations between this parameter and other model parameters.

Let's consider a scenario where all cities use the same $\omega$ value in the simulation. In this case, the sharing rate in large cities would be higher, while that in small cities would be lower. This theoretical sharing rate, determined solely by structural factors such as network structure, demand scale, and commute distances, can be regarded as the system's endogenous ride-pooling potential. However, the actual observed sharing rate is lower than this potential, especially in large cities, due to the suppressing effects of subjective resistance (e.g., higher $VoT$, higher $\omega$) and other unmodeled influences. Therefore, the purpose of calibrating $\omega$ is to adjust for the deviation of the actual willingness-to-share from this ideal value, and the calibrated $\omega$ value should be interpreted as a comprehensive reflection of all unmodeled influences, with $VoT$ being one of the important factors.

# Appendix I Pooling Ratio Loss

Different from the *shared passengers ratio* in the trip generation phase, the *pooling ratio* here is the result after the system completes all vehicle assignments and optimizations. Recalling Section 5.1.1, the proportion of shared passengers for each city was determined, which can be considered the theoretical upper limit of *shared passengers ratio* during the trip generation phase, based on actual observations from ODiN data and calibrated $\omega$ parameters. Typically, the finally realized *pooling ratio* will be lower than this theoretical upper limit from the trip generation phase.

This difference between the potential *shared passengers ratio* and the finally realized *pooling ratio* stems from several interconnected factors inherent in the operational reality. Primarily, vehicle availability and spatio-temporal constraints play a crucial role. Even if a potential shared trip meets the preferences of all participating passengers regarding $\omega$ and discounts, it might not be executed in the vehicle assignment phase because the situation of no suitable vehicle available at the specific time and location, or because dispatching any available vehicle would lead to pickup delays exceeding passenger tolerance. Furthermore, the system-level optimization, aiming to minimize the combined cost of passengers and the operator, introduces another criterion for selection. Shared trips that are highly attractive to passengers might still be rejected by the optimization algorithm if they significantly increase the overall system cost, due to factors like excessive detours and disruption to subsequent vehicle assignments. In such scenarios, solo trips that provide lower operational cost might be selected. Finally, the sequential assignment of vehicles based on timestamps introduces limitations. Decisions and vehicle assignments made for earlier trips reduce the availability of vehicles for subsequent trips. This means that the following shared trips might not be selected because a cost-effective vehicle cannot be found. In summary, these factors concerning vehicle dispatching, operational constraints, and the trade-offs within the system's optimization objective explain the difference between potential and realized sharing.

Table I.1: Descriptive Statistics of Potential and Realized Pooling Ratios

| **Statistic** | ***Shared Passengers Ratio*** | ***Pooling Ratio*** | ***Pooling Realization Loss*** |
|---|---|---|---|
| count | 37.000 | 37.000 | 37.000 |
| mean | 0.523 | 0.332 | 0.191 |
| std | 0.045 | 0.031 | 0.026 |
| min | 0.329 | 0.239 | 0.090 |
| 25% | 0.506 | 0.317 | 0.177 |
| 50% | 0.523 | 0.335 | 0.194 |
| 75% | 0.553 | 0.351 | 0.207 |
| max | 0.590 | 0.390 | 0.229 |

# Appendix J Calibration Details

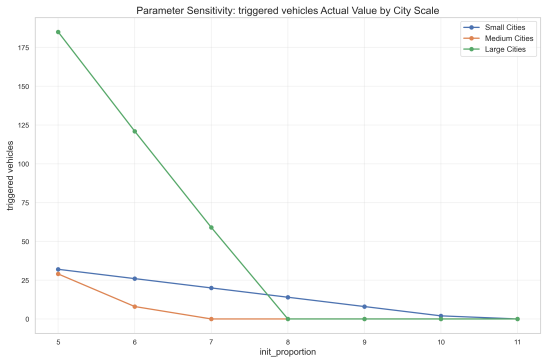
## J.1 Initial Proportion ($p_{init}$) Calibration

The parameter $p_{init}$ defines the calculation method for the initial fleet size, the useage of this parameter is by multiplying the total demand by this proportion factor, the result is the the number of vehicles initially deployed in the road metwork.

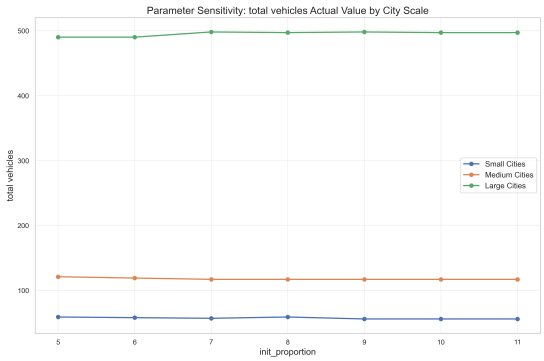
Although it has been observed that providing a larger initial fleet can slightly improve certain operational metrics (such as reducing average vehicle waiting time and total cost) by increasing the variety of vehicle options, it does not alter the final number of vehicles used by the system. Therefore, adopting the theoretical minimum value as the baseline offers the most transparent, logically clear, and representative starting point for all subsequent sensitivity analyses. This approach not only makes the results easier to interpret (reflecting the lower bound of necessary resources to serve all passengers) but also aligns with the practical consideration of resource constraints (the operators cannot have infinite number of SAVs) in real-world operations.

This study determines the default value of this parameter by analyzing its impact on the number of new vehicles triggered during the simulation process. The reason for choosing this metric to calibrate the initial fleet size parameter is the aim to minimize the introduction of new vehicles. As mentioned in Equation 3.8, the deployment cost calculation and pick-up time simulation for new vehicles are based on a fixed set of parameters. Under this situation, it is necessary for the system to eliminate the use of new vehicles to ensure the authenticity of the system results. Additionally, to ensure that the parameter analysis is conducted under the highest resource efficiency and clearly reflects the inherent characteristics of the strategy itself, it is needed to set the initial fleet size to a value that precisely generates this theoretical minimum fleet size. This is because, under the most constrained resource conditions, the system performance are more sensitive to parameters changes, while that a larger fleet could support the robustness of the system to certain fluctuations, making it less likely to observe subtle changes in system performance.

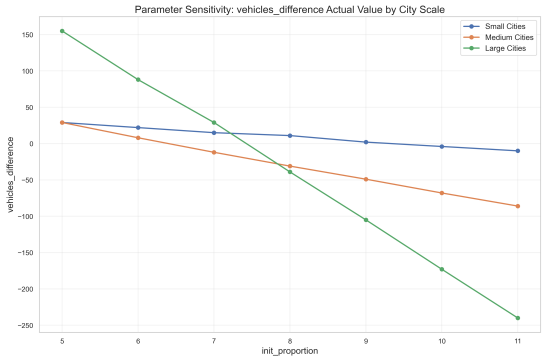
Thus, this study selects the smallest parameter value at which the number of additional vehicles triggered by the system equals zero. This ensures that no new vehicles are introduced, minimizing *bias* in the results, while also avoiding resource redundancy, which could similarly lead to biased outcomes. Here, the term *bias* refers to deviations from a peak performance pattern that was previously mentioned, which maximizes fleet resource utilization. To carefully determine a reasonably suitable initial fleet size, three representative cities from three different city scales are selected to conduct a sensitivity analysis on $p_{init}$ ranging from 6 to 11, with a step size of 1.



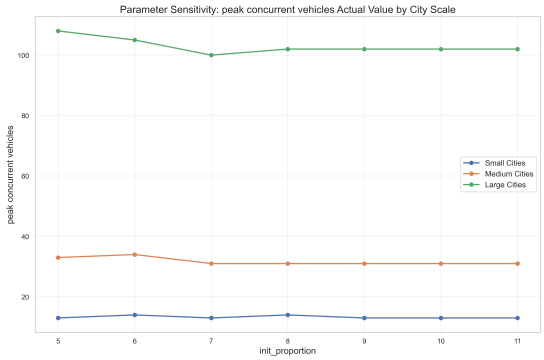

(a) $p_{init}$ vs. Number of Triggered New Vehicles




(b) $p_{init}$ vs. Total Number of Vehicles




(c) $p_{init}$ vs. Vehicle Difference




(d) $p_{init}$ vs. Peak Concurrent Vehicles

Figure J.1c clearly reveals the direct impact of the initial fleet proportion on the efficiency of vehicle resource allocation. It can be observed that for the three representative cities—Amsterdam (large city), Nijmegen (medium-sized city), and Zeist (small city)—the vehicle difference indicator turns negative after the initial fleet size surpasses a certain critical threshold. This phenomenon indicates that, at this point, the initially deployed vehicles exceed the actual operational demand, meaning no additional vehicles need to be introduced from external sources, and some of the initial vehicles may even remain unused throughout the simulation period, leading to potential idle resources in the fleet.

Further examination of Figure J.1b and Figure J.1d shows that when $p_{init}$ reaches and exceeds this critical threshold, both the total number of vehicles and the *peak concurrent vehicle* exhibit high stability, with their values hardly changing with further increases in the initial fleet size. Although the *peak concurrent vehicle* overlooks operational complexities such as spatio-temporal demand distribution, vehicle dispatching, and deadhead mileage, thus significantly underestimating the actual required fleet size, it highlights the substantial gap between the number of vehicles needed during actual operations and the core operational peak in simulations. This also validates the rationality of introducing the vehicle assignment mechanism. Compared to simply using the *peak concurrent vehicle* to estimate the required fleet size, the proposed method more accurately reflects the number of vehicles needed during real operations. Continuing to increase the initial deployed vehicles proportionally beyond this point does not further optimize these two core vehicle usage indicators; instead, it increases the number of unused initial vehicles, thereby reducing the overall operational efficiency and resource utilization rate of the fleet.

Therefore, the above analysis suggests that when determining the initial fleet size, a larger size is not necessarily better for operational efficiency. Once the initial fleet proportion exceeds a threshold sufficient to cover operational demand, the system no longer requires the activation of new vehicles, and both the total operational fleet size and the number of vehicles needed during peak periods stabilize at a consistent level. At this stage, further increasing the initial fleet size only results in more redundant resources, reinforcing the conclusion that the fleet resources are already adequate. This implies that when the initial fleet size is set according to this specific threshold, the system neither deploys excessive redundant initial vehicles nor struggles to meet demand. During subsequent simulation, the system can also manage with relatively few additional vehicles to satisfy full-day demand. If further reducing the initial fleet size, although the system might still maintain operations under ideal baseline demand, it would inevitably lead to more frequent triggering of the new vehicle introduction mechanism. This not only reduces the system's resilience in handling demand fluctuations and increases reliance on supplementary vehicles but

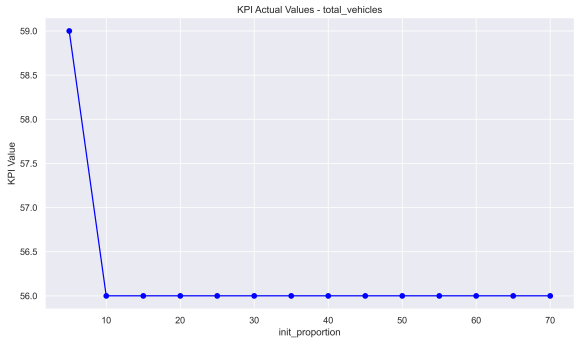
also, since the cost calculation and pick-up time simulation for newly introduced vehicles are based on a fixed set of preset parameters, imposes certain limitations on the generalizability and authenticity of simulation results across broader scenarios. From the results, it can be determined that the best inital fleet size for Amsterdam and Nijmegen using $p_{init}$ as 8, and Zeist as 11.

Due to the observation that there is a significant variation in the optimal $p_{init}$ across the three cities, to ensure that this study is able to consistently test the performance of different cities under the desired conditions, a set of suitable $p_{init}$ for all 37 cities are calibrated after analyzing these three cases. The calibration metrics remains the triggered new vehicles.

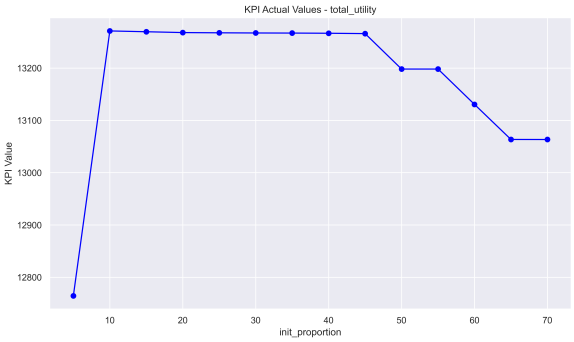
Table J.1: Optimal $p_{init}$ Statistics by City Size

| City Size | Mean | Std. Dev. | Count |
|---|---|---|---|
| Large | 8.38 | 0.51 | 8 |
| Medium | 9.11 | 2.02 | 19 |
| Small | 10.4 | 1.71 | 10 |

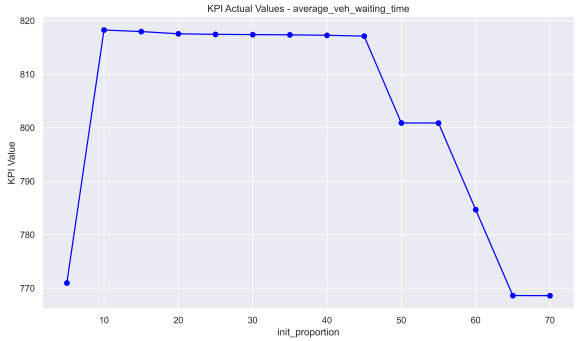
From the results, it can be concluded that, under the condition where full-day demand is set at 1%, small cities require a higher proportion of fleet size to meet the minimum level of fleet resource allocation. This trend is generally consistent with the patterns observed in the three representative cities that were tested earlier, given that medium-sized cities have the largest standard deviation. At this point, a key insight is that since the demand is positively correlated with population, if operators follow the strategy of deploying initial vehicles based on a certain proportion of demand, the minimum required fleet size for small cities often requires a higher proportion compared to larger cities.



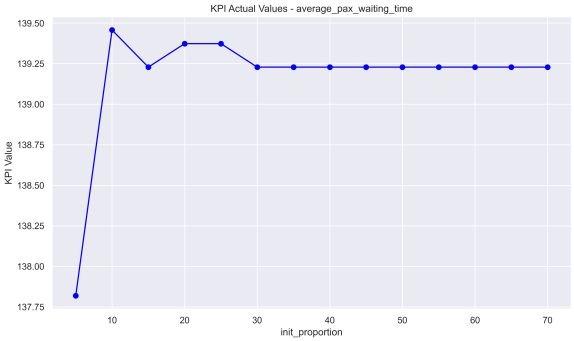

(a) Performance of Total Fleet Size




(b) Performance of Total Utility




(c) Performance of Average Vehicle Waiting Time




(d) Performance of Average Passenger Waiting Time

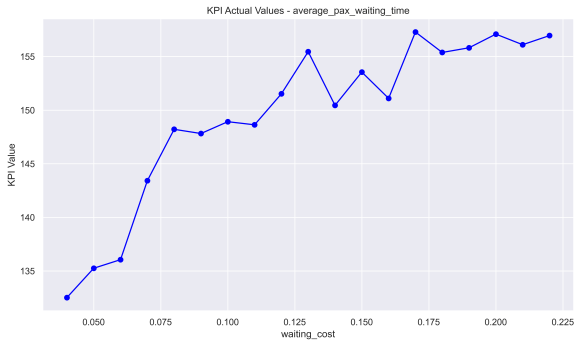
Figure J.2: Impact of Initial Fleet Size on Key Performance Indicators in Delft

Taking Delft as an example, a further analysis was conducted to test the system's performance in terms of the number of vehicles required and the composition of costs for various parties under conditions ranging from insufficient to extremely abundant vehicle resources. As shown in Figure J.2a, the number of vehicles peaks at $p_{init}$=5, which is due to the system lacking sufficient vehicles at this point, thus triggering the mechanism to introduce additional vehicles. In contrast, at other values, the fleet size remains stable at 56 vehicles. This number represents the actual fleet size the system ultimately uses to

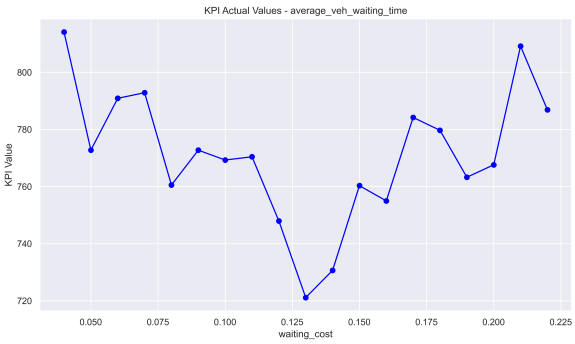
meet the demand of every passenger. Moreover, as vehicle resources become increasingly abundant, it can be observed that the total utility begins to decline after $p_{init}$ exceeds 45, accompanied by a reduction in the average vehicle waiting time. This indicates that under conditions of extremely abundant vehicle resources, the system gains access to more high-quality options, thereby optimizing the overall system utility.

## J.2 Waiting Cost Calibration ($c_w$)

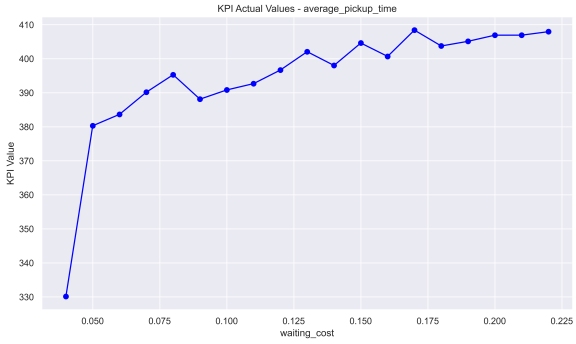
The parameter $c_w$ (€/second) serves as a penalty coefficient to quantify the cost incurred when a vehicle arrives at a passenger's pick-up point ahead of time and must wait. To clarify the significance of this parameter, although SAV do not have drivers and thus do not incur the traditional loss of driver time value, arriving too early still occupies potential parking resources. More importantly, it reflects inefficiencies in scheduling, as a vehicle waiting excessively could have been assigned to serve other more urgent or suitable trips. Therefore, introducing this penalty term aims to balance the idle waiting time caused by early arrivals. In previous experiments, $c_w$ penalty was set at 0.08, this was a value derived from simple preliminary tests designed to make the vehicle operating cost comparable to the vehicle waiting cost. To better understand the impact of $c_w$ on the system, a sensitivity analysis was conducted on this parameter to assess the system's response to its variation. Figure J.3 illustrates the effect of varying $c_w$ parameter within a focused range (0.04-0.22) with a step size of 0.01 on all the time-related metrics and vehicle reuse rates, using Rotterdam as an example.




(a) Rotterdam: Average Passenger Waiting Time




(b) Rotterdam: Average Vehicle Waiting Time

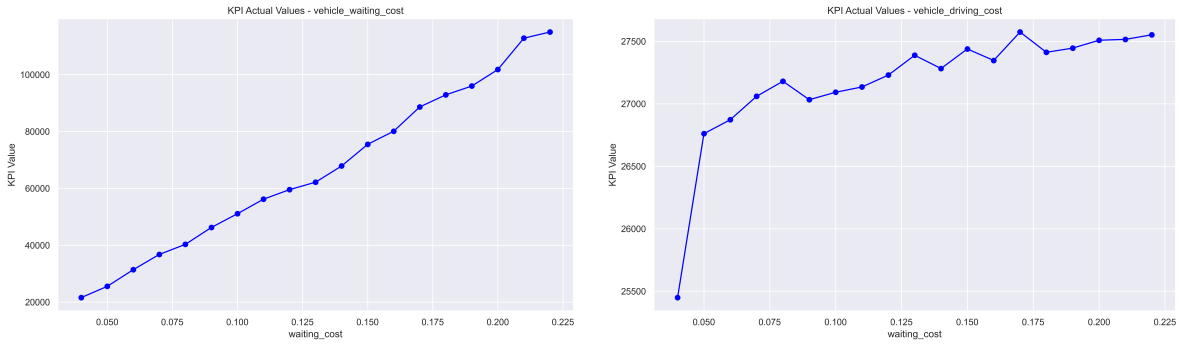



(c) Rotterdam: Average Pickup Time

Figure J.3: Impact of Different $c_w$ Values (0.04-0.22) on Time Metrics

The analysis indicates that the setting of $c_w$ has a complex impact on system behavior. As this cost coefficient increases, the system tends to reduce vehicle waiting time, but this is achieved by increasing pick-up time (dispatching farther vehicles that can depart sooner or new vehicles) or affecting passenger waiting time (potentially missing optimal matches). When $c_w$ is greater than 0.13, the system becomes extremely sensitive to waiting time, and the average vehicle waiting time unexpectedly increases, while passenger waiting time and pick-up time continue to rise slowly. This occurs because the system, being extremely sensitive to waiting time, attempts to assign a vehicle with the shortest possible waiting time for every ride. This can lead to the rejection of many nearby vehicle options that would require some

waiting, causing passengers to potentially wait longer to be matched with a vehicle that meets the zero-waiting requirement, thereby increasing average passenger waiting time and also raising travel costs and pick-up times.




(a) Vehicle Waiting Cost

(b) Vehicle Driving Cost

Figure J.4: Impact of Different $c_w$ Values (0.04-0.22) on Fleet Utility

To demonstrate the impact of the waiting penalty on utility, it can be seen from Figure J.4a and J.4b that when $c_w$ is greater than 0.05, the ratio of vehicle waiting cost to travel cost begins to increase. This suggests that the impact of vehicle waiting time on total utility starts to outweigh that of travel cost. Therefore, a balance point is needed to ensure that vehicle waiting time does not negatively affect total utility.

Considering this non-linear response, it is needed to select a parameter value that effectively utilizes the positive effects of the waiting cost penalty while avoiding pushing the system into extremely high-cost regions where efficiency might decrease. This desired value should result in a significant reduction in average vehicle waiting time compared to lower cost values, achieving the primary objective of introducing this penalty term. At the same time, average waiting time and average pick-up time remain at relatively reasonable levels, without significant worsening due to excessive avoidance of vehicle waiting. Furthermore, the vehicle waiting cost and driving cost become comparable in scale, which contributes to the balance of the objective function.

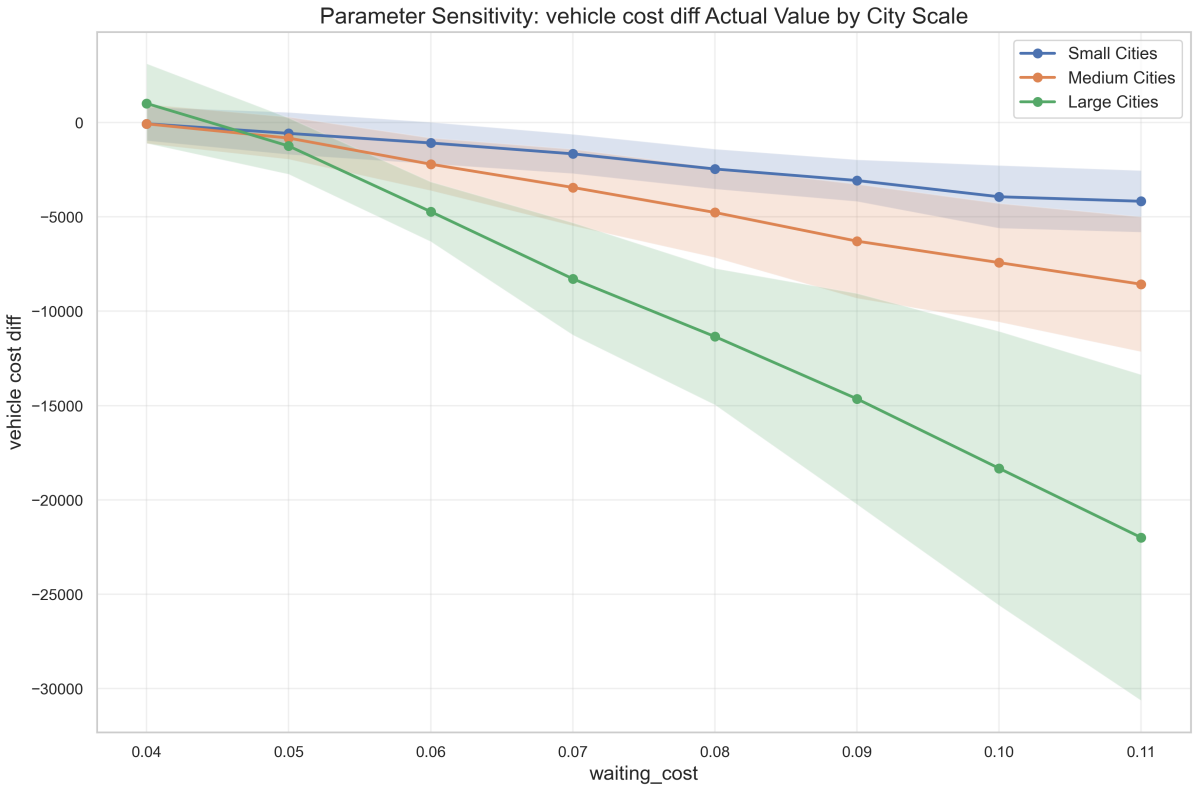
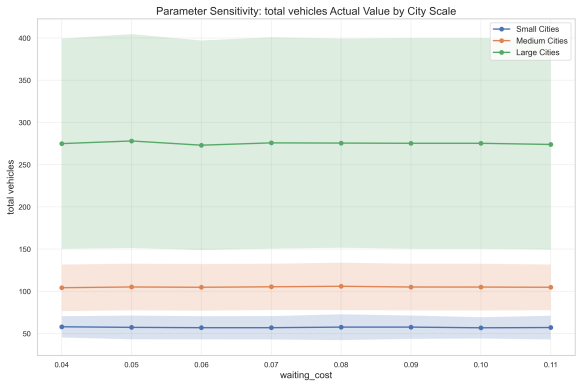


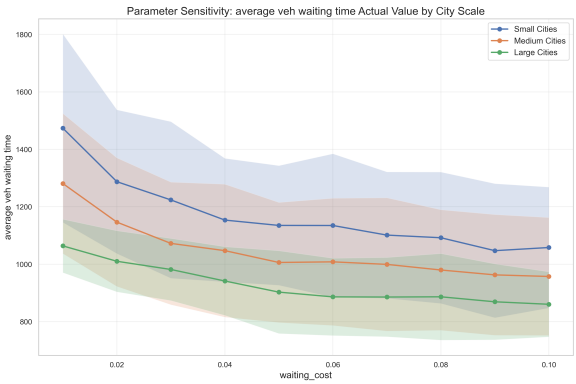

Figure J.5: Difference Between Vehicle Driving Cost and Vehicle Waiting Cost

A calibration test was conducted across all cities to observe the difference between vehicle driving cost and vehicle waiting cost. A negative value indicates that the vehicle waiting cost has exceeded the vehicle driving cost. As shown in Figure J.5, the balance point between vehicle driving and waiting costs for most cities is around 0.05. It can be seen that the previously setting of 0.08 makes the system excessively sensitive to vehicle waiting time, particularly for large cities. Based on this observation, *waiting cost = 0.05* is determined as the uniform default value in this study. Within the tested range, *waiting cost = 0.05* provides a well-balanced point. Furthermore, this value also makes the vehicle waiting cost and travel cost comparable in scale, which helps to balance the objective function. Therefore, selecting 0.05 as the default value aims to effectively suppress vehicle idle waiting while avoiding potential negative chain effects caused by setting the parameter too high, ensuring that passenger service levels and overall system operational efficiency are maintained at a good state.

In the meanwhile, to test the robustness of the previously determined initial fleet size, the performance of total fleet size and average vehicle waiting time under different $c_w$ values were tested. Figure J.6a and Figure J.6b show that changing $c_w$ here does not significantly affect the fleet size or the average vehicle waiting time of the fleet. This indicates that a relatively stable vehicle waiting time has been achieved by adding the waiting time penalty, further demonstrating the rationality of $p_{init}$ parameter calibration method and reflecting that the system's response to changes in $c_w$ parameter tends to stabilize.




(a) Performance of Total Fleet Size with Varying Waiting Cost




(b) Performance of Average Vehicle Waiting Time with Varying Waiting Cost

Figure J.6: Impact of Waiting Cost on Fleet Performance Metrics

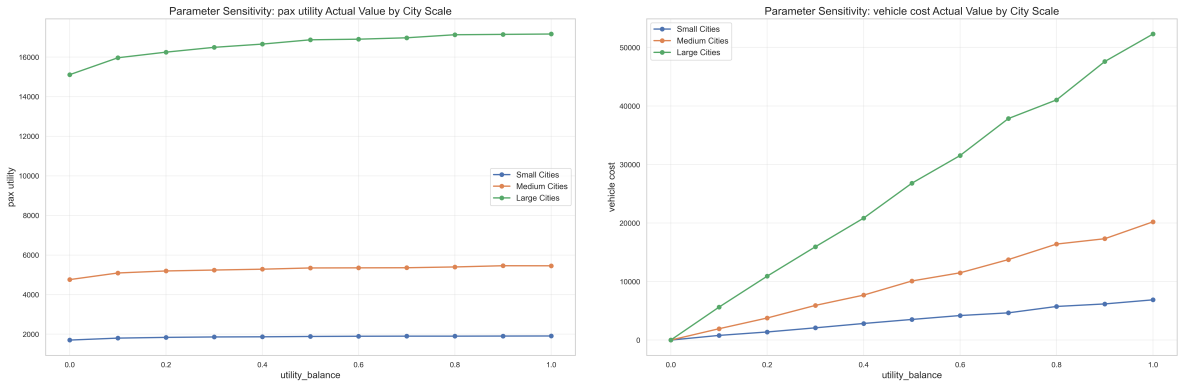
## J.3 Impact of Utility Balance ($\alpha$ )

Recalling the research objectives, to effectively evaluate and optimize SAV systems, this study sets the core goal as minimizing the total cost of the system. This choice is based on an understanding of sustainable SAV service operation models: a successful system must concurrently address operational efficiency and user attractiveness(Greifenstein, 2024). Therefore, the definition of total cost is set as a comprehensive metric that inherently includes both dimensions. These components have been converted into a unified monetary unit (Euro) through the theory of time value, enabling dimensional consistency and comparability.

The main reason for choosing to minimize this combined cost as the optimization objective is its ability to simulate and explore strategic balance points for achieving sustainable SAV services. On one hand, solely minimizing operational costs might lead to poor service quality, failing to attract and retain users; on the other hand, solely minimizing passengers' cost could result in excessively high operational costs, rendering the service economically unviable. Combining both into a unified objective function compels the optimization process to seek a trade-off solution that reduces operational costs while maintaining sufficient user satisfaction. This reflects the strategic decisions faced by operators in the real world: how to choose between cost control and service quality to achieve long-term market survival and development. Thus, the design of this objective function is not intended to precisely replicate the decision-making of any single party, but rather to find a balance point where the operator and passengers can mutually constrain each other, so that system optimization does not bias resources towards any one side.

However, simply adding both components with equal weight in the earlier judgement of this research is insufficient, because their economic scale and how they react to system changes might be different. This introduces the necessity of $\alpha$ parameter. This parameter, acting as a weighting factor, directly adjusts the relative importance of passenger utility and operational cost in the objective function, allowing for operationally explore different strategic balance points. Determining an appropriate $\alpha$ value is an essential step for all subsequent baseline performance evaluations and cross-scenario comparisons, as it ensures that the research is consistently conducted under a clearly defined optimization preference that aligns with the strategic goals. The objective is to identify a representative balance point where the marginal benefit of further adjusting $\alpha$ to optimize a specific target (e.g., reducing vehicle waiting costs) begins to decrease significantly. Therefore, it is necessary to conduct detailed simulation to determine a $\alpha$ value that better represents a reasonable strategic trade-off.

This experiment uses a two-stage sensitivity analysis. The first stage is to test different $\alpha$ values in three representative cities in different scales(Rotterdam, Nijmegen, and Zeist) to understand its impact and to find critical ranges. Then, the second stage is to repeat the analysis for all interested cities within the scope to confirm the results are valid and set the default value. It is not the goal to find an exact numerical match, instead, the focus is on observing the trade-offs between system performance (especially passenger utility and vehicle costs) using different weights. This helps identifing and selecting a $\alpha$ value that represents a reasonable strategic balance point. This value will be used as the default setting for later research.



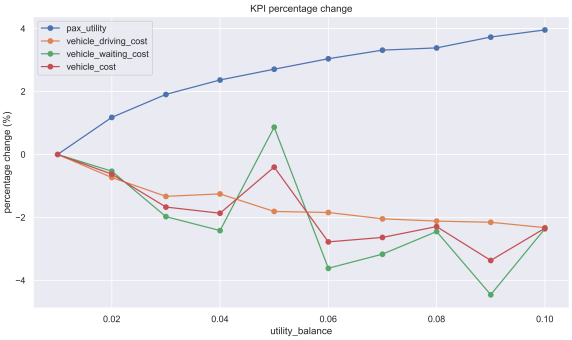

(a) Performance of Passenger Utility with Varying Utility Balance

(b) Performance of Operator Cost with Varying Utility Balance

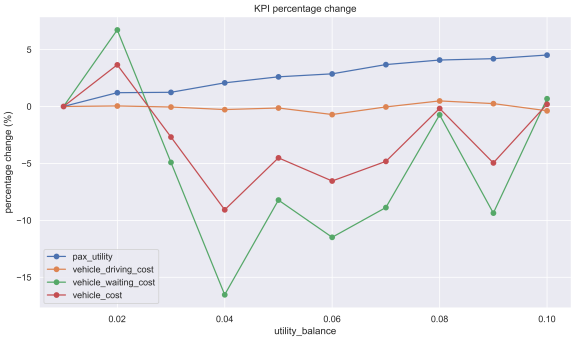
Figure J.7: Impact of Utility Balance on Performance Metrics

It can be observed that, due to the weighting by the utility factor, the fleet cost increases almost linearly. However, passenger cost only shows an improvement in the 0-0.1 range. This mainly indicates that passenger utility is fixed to a limited range. This is, in fact, related to the proposed algorithm structure itself. The changable component of passenger cost lies only in the passenger waiting time; other components are determined during the feasible rides generation phase(ExMAS) and the final ILP phase. Even if the system employs more ride-sharing to reduce some costs, because the ride-sharing rate itself is limited, the cost reduction from ride-sharing is also limited, so the optimization potential in this area is not high for users. This reflects that when operators anticipate passenger demand, they can already reduce passenger costs to a near-optimal level. However, the cost of the operators largely depends on the vehicle allocation process, as vehicles incur costs not only when serving passengers but also when waiting for and traveling to passenger locations. Therefore, fleet costs have more potential for variation in this scenario.

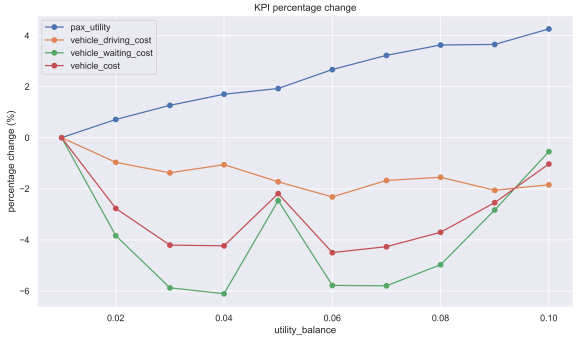
Consequently, from the perspective of the research objectives, simply assigning equal weights to operator utility and passenger utility is unreasonable, as this would cause the system to focus excessively on fleet cost while neglecting passenger utility. In the following simulation, the utility factor was removed when calculating fleet-related cost metrics. It is worth mentioning that this action does not affect the cumulated operator cost during ILP optimization phase, the intention is to eliminate the effect of the utility factor when generating KPIs to obtain a more intuitive understanding of the changes in fleet costs. The sensitivity of $\alpha$ was tested between 0 and 0.1, and the results are shown in Figures J.8a, J.8b, and J.8c.




(a) Rotterdam: Performance of Costs for Different Parties



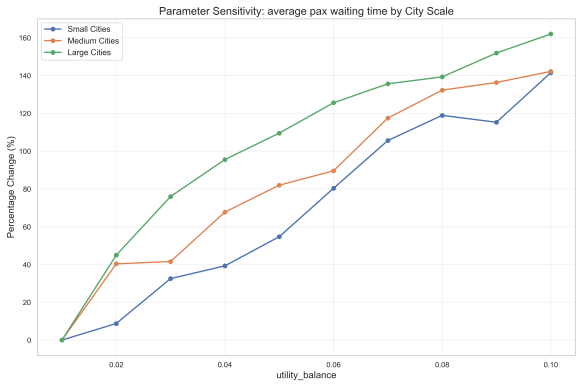

(b) Nijmegen: Performance of Costs for Different Parties



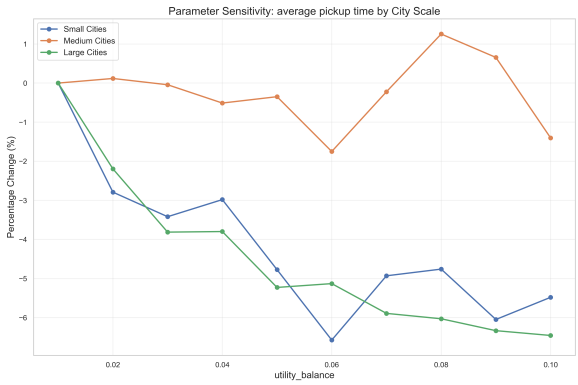

(c) Zeist: Performance of Costs for Different Parties

Figure J.8: Performance of Costs for Different Parties during Utility Balance Sensitivity Analysis

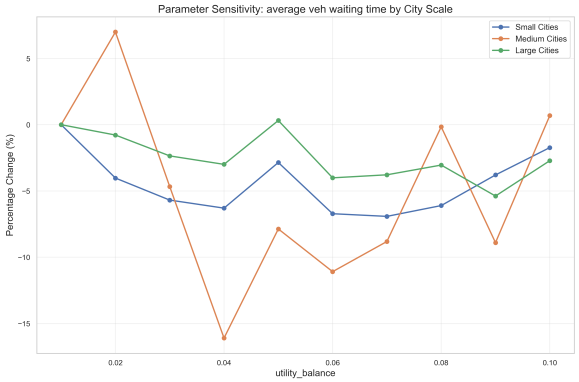
A clear trade-off relationship is shown in the figures. Increasing the weight of vehicle cost in the optimization objective indeed leads to a significant reduction in the actual total vehicle operation cost, with the maximum reduction ranging approximately from 4-15%. The cost breakdown, as shown in Figure J.9, indicates that this cost saving mainly stems from a substantial decrease in vehicle waiting time and vehicle pick-up time. However, this improvement in operational efficiency comes at a cost. It can be observed that passenger cost has increased by approximately 4-5%, which corresponds to the significant increase in passenger waiting time observed in Figure J.9.




(a) Passenger Waiting Time



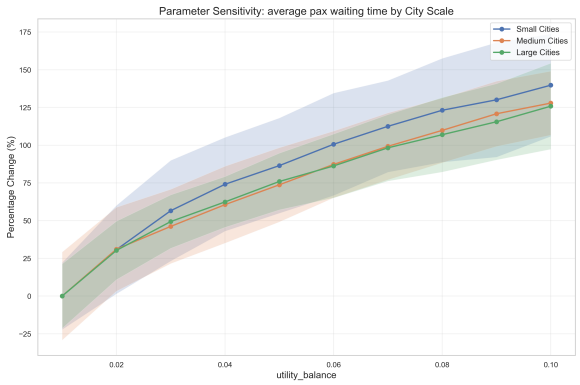

(b) Vehicle Pickup Time



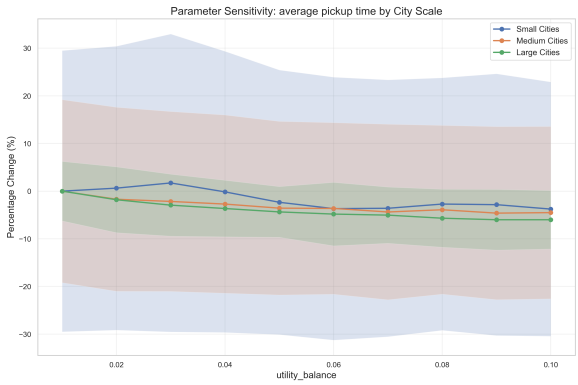

(c) Vehicle Waiting Time

Figure J.9: Impact of $\alpha$ on Various Time Metrics

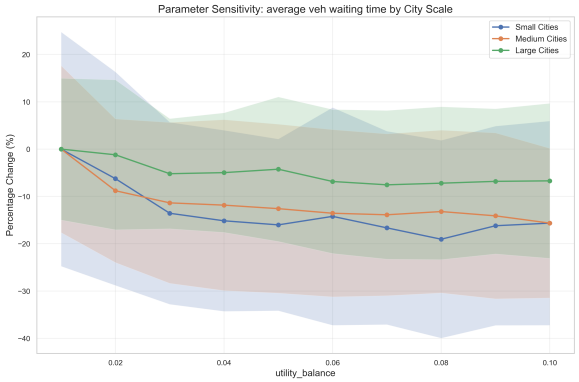
Previous observation for all cities shows that the cost-reduction potential of $\alpha$ is limited when it is greater than 0.1. Therefore, a further refined analysis was conducted by identifying an turning point before 0.1, aiming for a result closer to the point of diminishing marginal returns. However, due to urban heterogeneity, as shown in Figure J.9, a widely varying sensitivities to the changes in $\alpha$was found for different cities. Also, outliers are observed due to system complexity, making it difficult to determine the exactly value that this optimal turning point lies. A further analysis for all cities shows that even though the turning points for some cities are clear, finding a universal calibration method for all cities is extremely challenging. Further calibration would then deviate from the initial goal of analyzing fleet operation strategies across different city scales from a macroscopic perspective. This is because, at present, drawing a statistically significant conclusions regarding the relationship between $\alpha$ and city scale is infeasible. Therefore, a general analysis of $\alpha$ for all cities is conducted to find the turning point values for all cities on average. As shown in Figure J.10, it can be confirmed that 0.06 is a reasonable value. At this point, the rate of decrease in fleet waiting cost declines, and the increase in passenger cost also reaches a turning point, tending to stay at a stable level.




(a) Passenger Waiting Time




(b) Vehicle Pickup Time




(c) Vehicle Waiting Time

Figure J.10: Impact of $\alpha$ on Various Time Metrics (All Cities)

From the previous analysis, a balance point where vehicle and passenger interests can mutually constrain each other has been defined. At the same time, the significant heterogeneity among cities, as observed in the figures, indicates that the outliers and anomalies that were previously noted in the three chosen cities are not isolated cases. Even if it is possible to identify the turning point of diminishing marginal returns for each city, it is currently doubtful whether a statistically significant relationship could be established between these individual turning point values and city scale or other indicators. This further eliminates the necessity of pursuing individually optimal parameters.

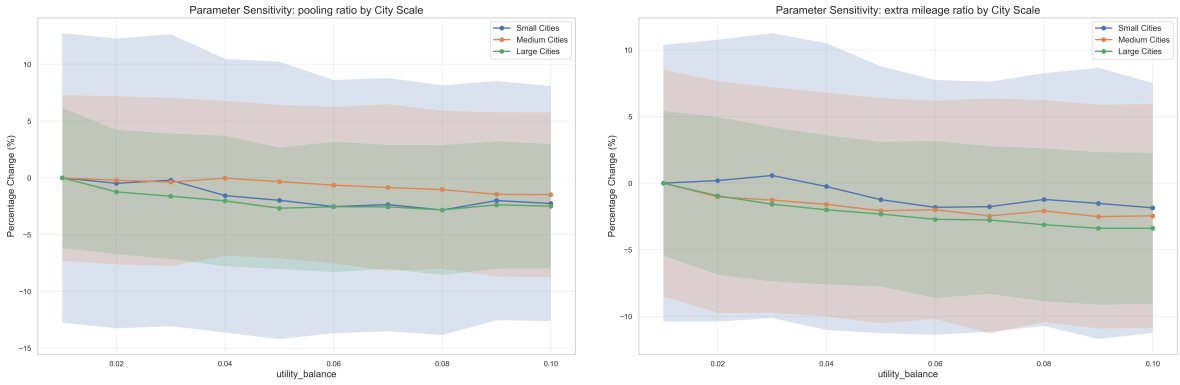
The sensitivity analysis of the utility balance parameter reveals several key strategic implications for the operation of SAV systems. This parameter corresponds to the extent to which policies aimed at improving user welfare or reducing fleet operational costs should be prioritized by operators.

The model clearly demonstrates that excessively prioritizing either party can lead to inefficient overall system performance or unsustainability. Operators can neither sacrifice essential service levels for unlimited cost reduction nor ignore economic feasibility in pursuit of the ideal user experience; they must consciously set goals and seek an effective balance point. Furthermore, the previous analysis points to an operational range with relatively optimal cost-effectiveness, represented by a turning point around $\alpha \approx 0.06$. Since all relevant costs are represented in monetary form, it might seem that this parameter excessively neglects the actual costs paid by operators. However, this is based on the results of a one-day simulation. In the long run, operator costs may reach a lower value after diminishing returns to scale, at which point the meaning expressed by this parameter will become more realistic.

Before this weighting factor reaches 0.06, by moderately increasing the focus on optimizing vehicle costs, significant improvements in vehicle efficiency (mainly by reducing waiting costs) can be achieved at a relatively small cost to passenger experience. However, outside this region, the marginal benefits of further strengthening vehicle cost optimization begin to decrease significantly, while the negative impact on passenger experience may continue to accumulate. In some cities, a rebound in total vehicle costs arises due to excessive compression of waiting times, leading to decreased travel efficiency. This suggests that operators should strive to adjust their operational strategies to be near this range, avoiding the blind pursuit of extremes in a single metric, with *ub=0.06* providing a data-supported reference point. Simultaneously, the study also finds that ensuring a baseline service level satisfactory to users is relatively

easy to achieve in the model (passenger costs tend to saturate even at low $\alpha$ values), whereas operational costs (especially vehicle waiting costs) exhibit high sensitivity to policy adjustments and offer greater optimization potential. Therefore, after ensuring basic service quality, operators should focus their main optimization efforts on improving vehicle resource management efficiency, reducing vehicle waiting time, and related operational costs, as this area has greater potential and a more significant impact on overall economic viability.

However, the formulation of operational strategies must also consider the significant impact of city characteristics. The high variance that was observed and the differences in response patterns among cities of different scales indicate that the idealized optimal balance point may vary depending on the city, and there is no absolutely universal optimal parameter. Nevertheless, in the absence of ideal conditions for fine-grained calibration for each city, the baseline value determined based on average trends still provides a valuable starting point and reference for the development of general strategies.




(a) Performance of Pooling Ratio with Varying Utility Balance

(b) Performance of Vehicle Extra Mileage Ratio with Varying Utility Balance

Figure J.11: Impact of Utility Balance on Performance Metrics

An analysis on the relationship between $\alpha$ and the pooling ratio is also needed since the pooling ratio is a crucial indicator in this reseaarch. As shown in Figure J.11a, it can be seen that as $\alpha$ increases, the pooling ratio slightly decreases. This phenomenon can be explained from two perspectives. The first direction is, why does a higher $\alpha$ lead to a lower pooling ratio? When selecting the optimal vehicle, since the vehicle's service time is dependent on the ride duration, the variable costs for the system mainly relies on the waiting time for early arrival and the pickup time from the vehicle's initial position to the passenger's location. When the fleet cost carries a higher weight in the total utility function, the system's optimization objective tends to allocate resources towards minimizing the vehicle's operational costs. Since pooled trips mostly occur during peak periods (when the number of concurrent trips is high), for example, if passenger 1 has potential pair-wise pooled trip options simultaneously with passenger 2 and passenger 3, then when the system iterate through rides by timestamp to match vehicles, there will be three ride allocation requests at that time stamp: passenger 1 traveling alone, passenger 1 pooling with passenger 2, and passenger 1 pooling with passenger 3. At this point, the system will assign three optimal vehicles as candidates for passenger 1's three options. It is almost certain to say that not all pooled trip options will be assigned a better vehicle compared to other options. Therefore, for pooled rides during peak trip periods, finding a vehicle with a short pickup distance and short waiting time becomes more difficult. Conversely, for a solo trip at a single timestamp, it is easier to find a vehicle that can respond quickly and has both short pickup and waiting times, thus achieving lower vehicle costs, reflecting its scheduling flexibility. Second, even though pooling can theoretically distribute certain fixed fleet costs (such as vehicle pickup time and vehicle waiting time), because all passengers share the same delay utility penalty, the advantage of cost distribution is not significant if the vehicle arrives late. Therefore, under a high vehicle cost weight, the optimizer tends to select the easier-to-implement solo trip options rather than executing pooled trip plans. The fundamental reason is that pooling often occurs during peak periods when vehicle resources are constrained, leading to potentially longer passenger waiting delays (see Figure 5.3), and this passenger waiting delay is effectively magnified for pooled trips. Thus, at such times, the system has greater flexibility in scheduling vehicles for each solo trip compared to pooled trips.

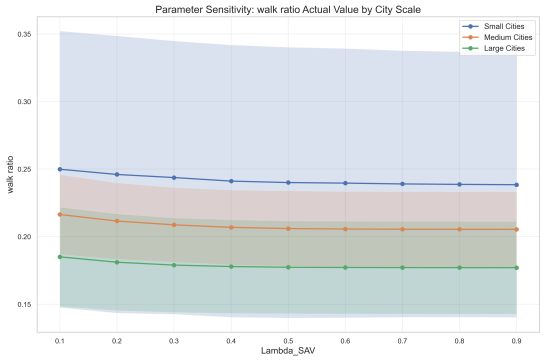
The second perspective addresses why a lower $\alpha$ leads to a higher pooling ratio. This occurs because when $\alpha$ is lower, the system optimization focuses more on minimizing passenger costs. In this study, the

system will try to arrange a specific passenger to a shared ride only if the pooling option is preferred or acceptable for the passenger in the perspectice of utility. The question will then focus on, why this process leads to an increase in various fleet costs? Figure J.11b shows that the extra mileage ratio increases when the passenger weight is higher. Recalling Equation 3.7, the primary ways for passengers to reduce costs in the system are by minimizing waiting time and increasing the pooling rate. Therefore, it can be inferred that, to reduce passenger waiting time, the system tends to assign vehicles that finish their previous service earlier with longer pickup distances, but can arrive at the passenger's pickup location earlier. This is because any user waiting is considered unacceptable under this parameter setting, making such a system behavior inevitable. Furthermore, to increase the passenger pooling rate, the system tends to assign vehicles that may not be optimal in terms of operator costs during peak hours to serve trips. Consequently, the operator's costs increase, while the passenger pooling rate improves.

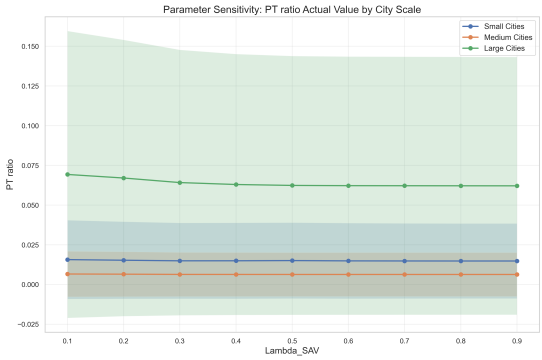
In the perspective of the operators, finding a vehicle with a short pickup distance and minimal waiting time for a solo trip is generally easier than finding an equally low-cost vehicle for a complex pooled ride. Therefore, if the operator's goal is to alleviate congestion or enhance vehicle utilization by enhancing the pooling rate, it might be necessary to dispatch vehicles that can respond more quickly to passenger requests, thereby reducing passenger waiting times. This may require accepting some vehicle assignments that are not optimal from the operator's perspective to facilitate pooling. Consequently, a careful assessment is needed to determine whether the societal and operational benefits derived from the increased pooling rate can compensate for the associated increase in operational costs.

# Appendix K Robustness Analysis of $\lambda_{SAV}$

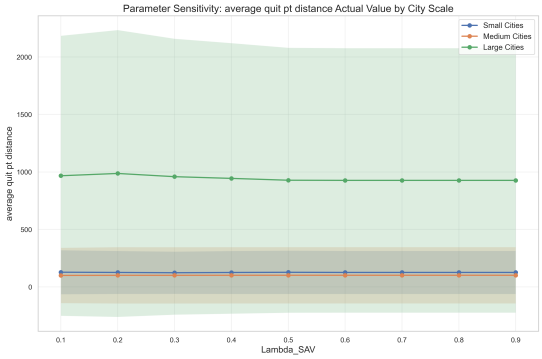
This section performs a robustness analysis on $\lambda_{SAV}$, a key parameter within the nested Logit model. This parameter, ranges from 0 to 1, governs the degree of substitutability between solo rides and shared rides for passengers within the SAV nest. A lower value of $\lambda_{SAV}$ implies that passengers are highly sensitive to the utility differences among the in-nest options, whereas a higher value indicates that choices become more stochastic. The experiment is designed to determine the extent to which the assumption of substitutability between solo rides and shared rides within the SAV nest, controlled by the $\lambda_{SAV}$ parameter, affects the model's predictions of the macroscopic competition structure between SAV and PT, thereby assessing the robustness of the core findings.




(a) Share of walking trips




(b) PT mode share




(c) Avg. PT distance of diverted trips

Figure K.1: Sensitivity Analysis for the $\lambda_{SAV}$ Parameter

Within the nested Logit model framework, the $\lambda_{SAV}$ parameter reflects the degree of substitution among the alternatives within a nest. As $\lambda_{SAV} \to 0$, it indicates that the in-nest alternatives are highly correlated, and passengers perceive them as nearly identical. In this case, the composite utility of the nest approaches the utility of the alternative with the highest utility within that nest. As the value of $\lambda_{SAV}$ increases, the independence of the in-nest alternatives strengthens. When $\lambda_{SAV} = 1$, the nested Logit model degenerates into a standard multinomial Logit model, where all alternatives are considered completely independent. Therefore, the value of $\lambda_{SAV}$ determines the extent to which the lower-utility(higher cost) solo ride option can enhance the overall attractiveness of the SAV nest.

As shown in FigureK.1, as $\lambda_{SAV}$ increases from 0.1 to 0.4, both the PT mode share (FigureK.1b) and the average PT distance of trips diverted to PT (FigureK.1c) exhibit a slight decrease, stabilizing after $\lambda_{SAV} \geq 0.4$.

This trend is consistent with the model's theory. When $\lambda_{SAV}$ is low, the composite utility of the SAV nest is primarily determined by the superior shared ride option, while the lower-utility solo ride option contributes minimally. This results in a lower overall attractiveness for the SAV nest, causing some passengers to shift to PT. As $\lambda_{SAV}$ increases, the utility of the solo ride is given greater weight in the calculation of the SAV nest's composite utility, which enhances the nest's overall competitiveness. Consequently, it attracts some passengers away from PT, leading to a decrease in the PT mode share.

The model results exhibit stability for $\lambda_{SAV} \geq 0.4$. This phenomenon is directly linked to the model's

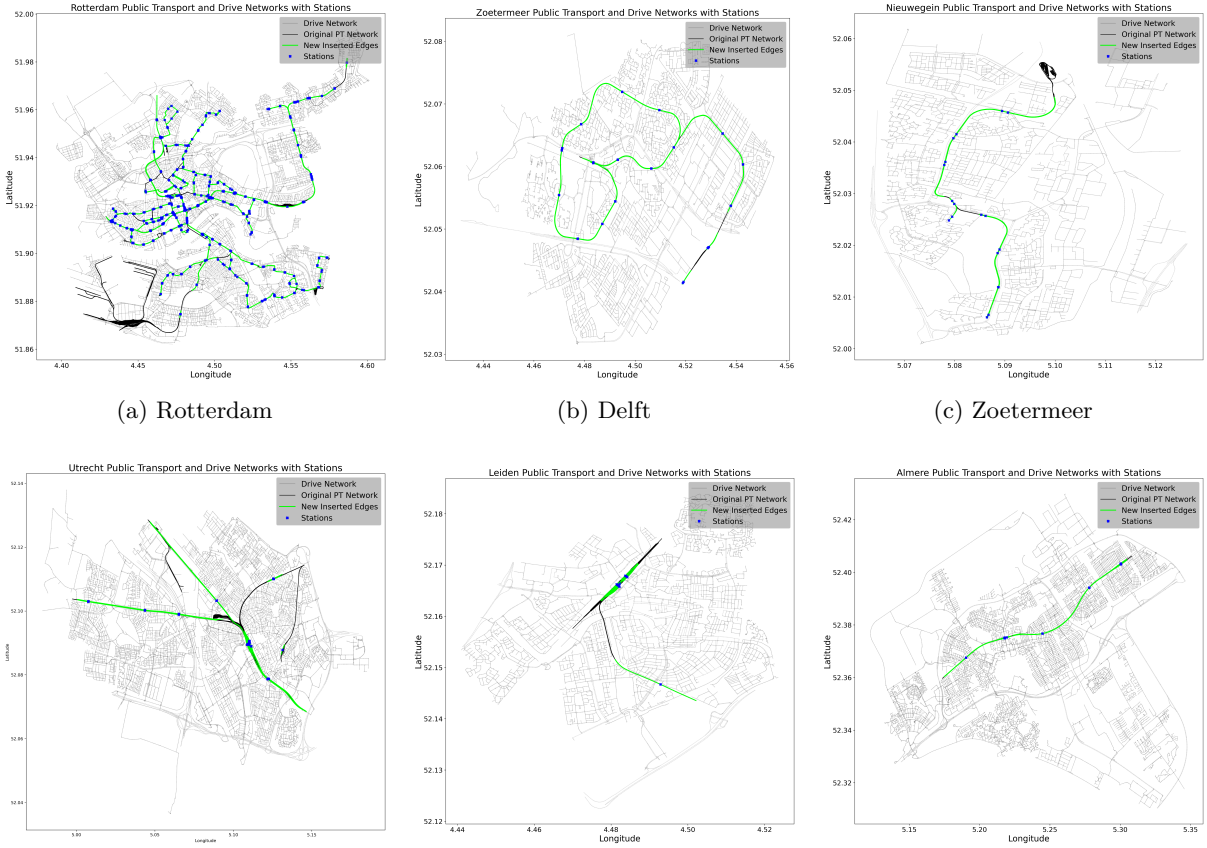
internal parameter configuration. According to the utility function defined in this study, the price of an solo ride is set higher than that of a shared ride. Meanwhile, the additional time penalty for the shared mode is, in most travel scenarios, insufficient to offset its price advantage. This setup creates a systematic and inherent utility gap between the two SAV modes. Consequently, when the value of $\lambda_{SAV}$ increases to a point where the utility contribution from this suboptimal solo ride alternative becomes saturated, its marginal enhancement to the overall utility of the SAV nest diminishes, which in turn leads to the stabilization of the model's macroscopic outputs.

In summary, this sensitivity analysis demonstrates that the model's response to variations in the $\lambda_{SAV}$ parameter is fully consistent with choice theory. Rather than merely validating a single parameter value, this analysis serves as a guide to how different assumptions about user behavior influence the predicted competition structure.

Specifically, if a lower $\lambda_{SAV}$ is assumed (e.g., $< 0.4$), suggesting a high correlation between SAV options, the model predicts a slightly reduced overall competitiveness for the SAV nest, leading to a marginally higher market share for PT. Conversely, assuming a higher $\lambda_{SAV}$ (e.g., $\geq 0.4$) implies greater independence between the SAV modes, which results in a more stable and slightly stronger market position for SAVs.

# Appendix L PT Network Statistics

Since public transport stations are not located directly on the nodes of the public transport network, auxiliary edges(green lines) are inserted to connect each station to its nearest point on the network. This integration ensures that stations become accessible nodes within the multimodal network, enabling accurate and realistic path computation across different transport modes.




(a) Rotterdam (b) Delft (c) Zoetermeer

(d) Utrecht (e) Leiden (f) Almere

Figure L.1: Comparison of rail public transport network structures: effective competition group (top) vs. network-restricted group (bottom).

Figure L.1 compares the public transport network structures of cities in the effective competition group (top row) and the network-restricted group (bottom row). In each row, the cities are arranged from left to right as large, medium, and small, respectively. As illustrated by the comparison between Rotterdam and Utrecht—both classified as large cities—Rotterdam's PT network provides much higher accessibility than that of Utrecht. This difference in coverage fundamentally determines the competitive relationship between PT and SAV: higher PT coverage leads to stronger PT competitiveness. The significant difference in PT mode share between the two cities is mainly caused by the disparity in PT network accessibility.