TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Fan, Q., van Essen, J. T., & Homem de Almeida Correia, G. (2026). Solution methods for pricing and fleet management in shared automated vehicle services considering supply-demand dynamics, congestion, and income heterogeneity. *Transportation Research Part C: Emerging Technologies*, *194*, Article 106011. https://doi.org/10.1016/j.trc.2026.106011

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

*This work is downloaded from Delft University of Technology.*

Transportation Research Part C 194 (2027) 106011

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

# Solution methods for pricing and fleet management in shared automated vehicle services considering supply-demand dynamics, congestion, and income heterogeneity

Qiaochu Fan [a,*], J. Theresia van Essen [b], Gonçalo Homem de Almeida Correia [c,1]

[a] Department of Transport & Planning, Delft University of Technology, Delft, 2628 CN, the Netherlands
[b] Delft Institute of Applied Mathematics, Delft University of Technology, Delft, 2628 CD, the Netherlands
[c] CITUA, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, Lisboa, 1049-001, Portugal

ARTICLE INFO

*Keywords:*
Supply-demand interaction
Shared automated vehicles
Pricing
Fleet sizing
Fleet management
Traffic congestion

ABSTRACT

Research on the supply–demand dynamics of Shared Automated Vehicle (SAV) services has grown rapidly. However, the resulting models are often highly complex nonlinear systems that are difficult to solve. This study has two main objectives: (1) to investigate profit-maximising pricing and fleet management strategies for SAV services while accounting for supply–demand interactions, congestion effects, and travellers' heterogeneous income levels; and (2) to develop and compare several solution methods for determining these strategies. We formulate a mixed-integer nonlinear programming model for the SAV fleet management problem under three pricing strategies: a base fare plus a distance-based fare, a distance-based fare only, and an income class–based fare. To solve this model, we propose three algorithms designed to address its nonlinearities: a linearisation-based method (P1), a hybrid particle swarm optimisation method (P2), and a hybrid Bayesian optimisation method (P3). We also examine the feasibility of the solutions obtained by these methods with respect to the original problem, and introduce a correction procedure when infeasibility arises. The proposed methods are evaluated through a small case study and a larger case study based on Delft, the Netherlands. The results show that all three methods can find good-quality solutions within short computational times in the small case study, while P3 achieves the best objective value in the Delft case study. From a practical perspective, the Delft case study shows that the first pricing strategy can encourage active transport for short trips, the second leads to the lowest congestion but also the lowest profit, and the third achieves the highest profit by attracting more demand and exploiting the lower price sensitivity of high-income travellers, despite causing higher congestion. Additional analyses further examine the effects of income class segmentation, income distribution, and congestion.

## 1. Introduction

Supply-demand interactions in the ride-hailing market are an important element that should not be ignored when making policy, management and operational decisions. This is especially the case when it comes to researching future mobility systems, such as

* Corresponding author.
*E-mail addresses:* q.fan-1@tudelft.nl (Q. Fan), goncalo.h.a.correia@tecnico.ulisboa.pt (G. Homem de Almeida Correia).

[1] Given his/her/their role as <Editor>, <Gonçalo Correia> had no involvement in the peer-review of this article and has no access to information regarding its peer-review. Full responsibility for the editorial process for this article was delegated to another journal editor.

https://doi.org/10.1016/j.trc.2026.106011
Received 2 September 2026; Accepted 5 September 2026
Available online 22 September 2026
0968-090X/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Shared Automated Vehicle (SAV) services, where the future demand for SAVs is endogenously determined by pricing and system design decisions. Similar to conventional ride-hailing, users request trips on-demand, and the system matches users with available vehicles. However, unlike traditional services, SAVs are operated by a centralised fleet management system, which coordinates vehicle routing and repositioning in real time.

Such a future mobility system consists of a variety of components that depend on each other interactively. From the perspective of the SAV service suppliers, decisions need to be made at the planning and operational levels. Planning decisions, made before launching the SAV service, might include aspects like pricing strategy, SAV fleet sizing, initial fleet distribution, and service quality. Operational decisions, made in response to new SAV service requests and real-time network conditions, may involve trip assignment, vehicle routing, relocation, parking, and decisions to accept or reject requests. These decisions, in turn, influence passengers' choice to use the service.

On the demand side of the SAV service system, passenger choices are notably influenced by factors such as price, travel time, and service quality. Moreover, heterogeneous passengers with diverse socio-demographic characteristics, including variations in income level, age, and gender, demonstrate different sensitivities to these factors (Kolarova et al., 2019; Kolarova and Cherchi, 2021; Ma et al., 2023). In such an SAV service system, the decisions on the supply side are inherently linked to those on the demand side. However, many studies tend to assume that demand is predetermined when optimising supply-side decisions, and overlook the time-varying network conditions or fail to consider the heterogeneous preferences of passengers. This triggers the need for modelling the complex interactions of these decision variables dynamically. Developing such models is crucial for enabling service suppliers to make profitable and realistic decisions in the management and operations of current and future systems.

Among all the planning decisions, pricing exerts the most direct and significant influence on passengers' willingness to use the service (Chen et al., 2020). Pricing is an effective tool not only for balancing supply and demand but also for alleviating traffic congestion. Consequently, understanding demand patterns and designing effective pricing strategies has attracted considerable interest from both academia and the industry. Although various pricing strategies and fare structures have been explored, few studies have investigated optimal pricing strategies tailored to diverse demographic groups. In this paper, we investigate three pricing strategies: (1) a base fare plus a distance-based fare, (2) a solely distance-based fare, and (3) an income class-based fare. Income class-based fares devise prices according to passengers' income levels, making SAV services more accessible to low-income individuals and potentially boosting overall usage and social equity (Verbich and El-Geneidy, 2017; Dong et al., 2022). Moreover, charging higher fares to high-income users could balance the service's overall affordability and enhance sustainability.

In this paper, we investigate the optimal pricing and fleet management strategies for a SAV service provider operating a ride-hailing fleet in a future urban mobility system. Specifically, we focus on a private car-free urban area where road space is reserved for public transport and active modes of travel, aiming to mitigate challenges such as traffic congestion, air pollution, and accidents involving vulnerable road users (Nieuwenhuijsen and Khreis, 2016; Van Arem et al., 2022). This vision is increasingly reflected in urban policies worldwide, as many cities recognise that reducing or eliminating private car ownership can substantially enhance public health, sustainability, and urban liveability. Cities such as Hamburg, Oslo, Helsinki, and Madrid have introduced or planned car-free zones, while others, including Bogotá, Brussels, Chengdu, Copenhagen, and Paris, have implemented regular car-free days as transitional strategies (Nieuwenhuijsen and Khreis, 2016). In the Netherlands, seven cities (Amsterdam, Rotterdam, Den Haag, Utrecht, Eindhoven, Helmond and Groningen) are increasingly adopting goals for climate neutrality by 2030, with a strong interest in car-low zones as part of their urban planning strategies.

Within such contexts, SAVs have emerged as a promising complement or alternative to traditional public transport, offering flexible, on-demand, and efficient mobility services (Spieser et al., 2014; Liang et al., 2020). Previous studies show that SAVs can satisfy urban mobility demand with significantly fewer vehicles, thereby reducing congestion, emissions, and parking requirements (Fagnant and Kockelman, 2014; Spieser et al., 2014). In the envisioned car-free system considered in this study, travellers can choose between two main modes: SAVs and bicycles. Bicycles represent the active transport option, while walking is excluded because it mainly serves very short trips for which SAVs would not be a competitive alternative. A scenario with long-distance trips usually done by motorised modes falls outside the scope of the local demand-congestion interactions we model. Travellers use smartphone apps to request SAVs by entering trip details. The service platform evaluates these requests and accepts them if they are beneficial to the company. If a request is rejected, the traveller is assumed to choose the bicycle as the next-best alternative. Accepted requests are matched with available SAVs, which are dispatched to pick up customers.

The mathematical model proposed in this paper is a Mixed-integer Nonlinear Programming (MINLP) model adapted from the model by Fan et al. (2023), with two major modifications. Firstly, three pricing strategies are modelled additionally, with pricing decisions as endogenous variables. Secondly, the diverse sensitivities of different passenger income classes are incorporated into the demand modelling to explore their impact on these pricing strategies. In addition, the endogenous demand for SAVs is modelled using a binary logit model, where price, travel time, and service quality influence travellers' willingness to use SAV services. Traffic congestion is modelled by dynamically varying travel time based on vehicle flow on road links. The accept/reject decisions, which reflect service quality, are also endogenously determined in the model.

With the modifications to the model, its complexity has increased due to the non-linearity introduced by both non-linear constraints and a non-linear objective function, as well as the addition of new variables. These changes pose significant challenges for conventional solution methods. To address this complexity, the existing literature typically employs but is not limited to, the following strategies: (1) linearisation techniques (Guo et al., 2022; Fan et al., 2023), which convert non-linear components into a form that can be solved using linear programming, thereby simplifying the complexity of the original problem; (2) reformulation techniques (Huang et al., 2018; Lu et al., 2021), which reformulate the problem, identify problem structure or decompose the problem into smaller and more

manageable problems; (3) heuristic/metaheuristic algorithms that seek to find sufficiently good solutions for complex optimisation challenges where exact methods are impractical (Huang et al., 2018; Lu et al., 2021).

In response to these challenges, our research has developed three tailored solution methods. The first method utilises linearisation techniques. The second employs a hybrid approach that combines Particle Swarm Optimisation (PSO), a metaheuristic, with reformulation techniques. The third method employs Bayesian Optimisation (BO) alongside linearisation techniques. BO is particularly effective in contexts where evaluating the objective function is costly or complex, such as in parameter tuning for machine learning algorithms. In our study, it plays a key role in simplifying the objective function of the mathematical model. To enhance computational efficiency, the second and third methods are implemented in parallel. A comparative analysis of these methods is conducted to assess their effectiveness and efficiency in solving the proposed model. Effectiveness refers to the ability of a solution method to obtain high-quality solutions. Efficiency refers to the computational performance of a solution method, specifically how quickly it can produce a good-quality solution.

The main contributions of this paper are as follows:

(1) We develop an optimisation framework for SAV services that jointly determines pricing and fleet-management decisions under three pricing strategies. The framework explicitly captures demand–supply interactions, congestion effects, and income-based heterogeneity in travellers' preferences.
(2) We propose and compare three solution algorithms for the resulting MINLP problem, including a linearisation-based method, a hybrid PSO-based method, and a hybrid BO-based method. We further evaluate the feasibility of the obtained solutions with respect to the original problem and introduce a correction procedure to recover feasible solutions when needed.
(3) We conduct an extensive analysis of income heterogeneity, examining the impact of both income-class segmentation and variations in income distribution. The results demonstrate that accounting for income differences is essential for obtaining behaviourally realistic demand estimates, operationally sound fleet decisions, and policy-relevant pricing strategies.

The remainder of the paper is structured as follows. Section 2 reviews the literature on pricing strategies in ride-hailing services and on models that capture supply–demand interactions. Section 3 presents the nonlinear, non-convex mathematical model formulated in this study. Section 4 introduces the three proposed solution algorithms and explains each method in detail. We then assess the feasibility of the solutions obtained by the proposed methods with respect to the original model and presents a correction procedure for infeasible solutions. Section 5 illustrates the practical application of the proposed models through a small-scale case study and a quasi-real case study of Delft, the Netherlands. Finally, Section 6 concludes the paper by summarising the main findings and outlining directions for future research.

## 2. Literature review

The literature review is divided into two main sections: Section 2.1 focuses on the existing literature addressing pricing problems in ride-hailing services; and Section 2.2 dives into the literature which models the endogenous supply-demand interactions, presenting the demand modelling techniques and main methodologies. This review aims to provide a comprehensive understanding of both pricing problems and supply-demand modelling in the context of ride-hailing services.

### *2.1. Pricing problems in ride-hailing services*

Pricing problems in ride-hailing services have gained a wide range of research interests (Wang et al., 2016; Tong et al., 2018; Nourinejad and Ramezani, 2020; Zhang and Nie, 2021; Asadpour et al., 2023). Existing research on the practice of ride-hailing services focuses on a two-sided market, where both the fare paid by passengers and the wage paid to drivers need to be determined. However, in the context of using shared automated taxis for ride-hailing services, the problem is simplified because there is no need to hire human drivers anymore. Thus, we only focus on the pricing strategies for the fare paid by the passengers.

Pricing mechanisms can be categorised into two types: static pricing and dynamic pricing. Static pricing strategies determine a fixed price structure throughout the planning period. Research has investigated, amongst others, distance-based pricing (Dong et al., 2022), travel time-based pricing (Liang et al., 2020), origin- and/or destination-based pricing (Özkan, 2020; Müller et al., 2023), zone-based spatial pricing (Li et al., 2021; Dong et al., 2022). Dynamic pricing allows the price and/or wage to vary according to various factors such as the current market demand and supply relationship, time, or even competitor pricing. Examples of this include surge pricing (Al-Kanj et al., 2020), spatial-temporal pricing (Meskar et al., 2023), and congestion pricing (Zheng et al., 2023). Surge pricing mechanisms usually charge passengers a higher price during peak hours, extreme weather conditions, or major events where the supply-demand is imbalanced.

Existing research on static and dynamic pricing presents a range of controversial views. Nourinejad and Ramezani (2020) address three pricing strategies, which are static pricing, constrained dynamic pricing (assuming the instantaneous profit is also non-zero), and unconstrained dynamic pricing (allowing the instantaneous profit to be non-zero). Their results show that the unconstrained dynamic pricing strategy provides the highest overall profit and ensures a more stable rider waiting time with less variation during the study period. Similarly, Cachon et al. (2017) find that incorporating dynamic elements into pricing and/or wages can enhance profits compared to fixed pricing strategies. However, despite its profitability, surge pricing has faced criticism for its potential negative impact on consumer welfare, as it forces consumers to either pay higher prices or wait longer during peak demand periods (Ashkrof et al., 2022). Moreover, Lin and Zhou (2019) suggest that surge pricing is not always the optimal choice for ride-hailing

companies, as static pricing can achieve comparable performance. In this study, we focus on investigating static pricing strategies during surge demand periods (peak hours) in a typical working day. During this optimisation period, the pricing strategy stays the same.

People with different socio-demographic characteristics—such as gender, age and income levels—display different sensitivities to price changes. The behaviour of heterogeneous users has a great impact on the platform's optimal pricing and wage decisions (Taylor, 2018; Wu et al., 2020). Beirigo et al. (2022) segment users of an autonomous mobility-on-demand system into two classes: first-class users, who are willing to pay more for a higher level of service, and second-class users. Bai and Tang (2022) proposed a model to study the equilibrium between two competing on-demand service platforms. Their proposed model can capture some market characteristics such as the price- and time-sensitive customers and earning-sensitive service providers. Dong et al. (2022) introduced class-based pricing strategies in a chance-constrained dial-a-ride problem where heterogeneous users are grouped into classes by their socio-demographic characteristics. However, the pricing is introduced as a class-based parameter, which is not endogenously determined. In the existing pricing problems, very few have considered an income class-based fare, considering people's heterogeneous sensitivity toward the costs/money. Thus, in our work, we study the optimal income class-based fare and its interplay between this pricing variable and all the other planning and operational level variables.

### 2.2. Endogenous supply-demand interaction modelling

The supply-demand interaction has been studied in transportation systems for different problem settings, for example, in the two-sided market of ride-hailing services (Li et al., 2021; Meskar et al., 2023), carsharing systems (Huang et al., 2018; Lu et al., 2021), dial-a-ride problems (Dong et al., 2022), fleet sizing and management problems using SAVs (Guo et al., 2022; Fan et al., 2023), public transit planning problems (Cadarso et al., 2017; Steiner and Irnich, 2020; Wei et al., 2022), and competition among multiple on-demand mobility services (Wang et al., 2022). In Table 1, we list several papers that consider supply-demand interactions with their main decisions, methodologies, and how demand is modelled.

Accurately modelling the interactions between decision variables on both the supply and demand sides necessitates that demand is determined endogenously. Normally, demand can be modelled by (a) a simple linear or non-linear function such as an exponential function (Jorge et al., 2015; Huang et al., 2020); (b) a discrete choice model, such as a binary logit model (Lu et al., 2021; Guo et al., 2022; Fan et al., 2023), a multinomial logit model (Zhang and Nie, 2021), a nested logit model (Cadarso et al., 2017), a dynamic discrete choice model (Västberg et al., 2020), or chance-based constraints (Dong et al., 2022); (c) a simulation-based approximation of the discrete choice model (Paneque et al., 2021, 2022). The demand model is then integrated into optimisation models (Huang et al., 2018; Lu et al., 2021; Paneque et al., 2021; Zhang and Nie, 2021; Paneque et al., 2022; Fan et al., 2023), simulation-based methods (Hörl et al., 2021; Wang et al., 2022), or a hybrid of both (Pinto et al., 2020).

Incorporating demand modelling into an optimisation problem often presents challenges to the solution process, typically requiring the development of tailored solution algorithms. Some studies use aggregate equilibrium or macroscopic models to examine system-level interactions among ride-hailing, private traffic, and public transport. Fayed et al. (2023) investigate the use of dedicated bus lanes by pooled ride-hailing vehicles and show that the delay-minimising allocation may allow only a fraction of pooled vehicles to use these lanes, depending on demand levels across modes and the allocation of road space. Fayed et al. (2024) extend this line of research to a dynamic macroscopic setting, in which pooling fares relative to the solo fare are adjusted to steer the shares of pooled trips using the general-purpose and bus-lane networks. Such aggregate approaches provide tractable representations of system-level multimodal interactions, whereas our problem requires explicit trip assignment, vehicle routing, fleet sizing, and link-level time-varying congestion. Simulation-based methods excel at capturing microscopic details by reconstructing complex scenarios. However, they often suffer from long computational times, particularly when there are many decision variables to be combined. Therefore, they are often used to assess multiple scenarios instead of searching for an optimal combination. Different from most existing studies, we develop optimisation-based methods that not only describe the supply-demand endogenous interactions but also capture the detailed operational decisions of SAVs and the dynamic effects of network congestion. Due to the complexity of the problem, three tailored solution algorithms are designed to address the model efficiently.

## 3. Problem formulation

In this section, we first present the main assumptions in Section 3.1. Then, we describe the mathematical model, including model settings in Section 3.2, passenger demand modelling in Section 3.3, SAV service planning and operation in Section 3.4, traffic congestion modelling in Section 3.5, and the objective function in Section 3.6.

### 3.1. Assumptions

Some assumptions are made in this paper:

(a) It is assumed that the total number of trips within the urban area can be well estimated beforehand over the planning horizon. The number of trips and their origin and destination differ per time step within this planning horizon.
(b) The SAVs in this study operate at SAE level 5 (On-Road Automated Driving (ORAD) committee, 2021), allowing them to navigate the entire network autonomously without a human driver.
(c) Privately owned and human-driven vehicles are excluded from this study. Travellers can choose only between SAVs and bicycles. This framework can be easily extended to include additional modes, such as walking and metro.

**Table 1**
Selected papers on supply-demand interaction in transportation systems.

| Transport mode or service | Authors | Main decisions | Methodology | Demand modelling |
|---|---|---|---|---|
| Carsharing system competing with private cars | Huang et al. (2018) | Carsharing station location, station capacity, fleet size, vehicle relocation, and mode choice | A mixed-integer nonlinear programming model, solved by a customised gradient algorithm | Modelled using a binary logit model |
| Two-sided ride-sourcing market | Nourinejad and Ramezani (2020) | Passenger demand, drivers supply, fare for riders, wage for drivers, passenger waiting time, driver cruising time, | A macroscopic analytical model and a microscopic simulation model used within a model predictive control approach | Modelled as a function of generalised cost |
| Transit networks and shared autonomous mobility fleets | Pinto et al. (2020) | Transit frequency, fleet size of SAVs, mode choice, SAV route choice, traffic congestion | A bi-level mathematical programming formulation, solved by a heuristic procedure combining optimisation and simulation approaches | Modelled using a multinomial logit model |
| Carsharing system competing with private cars | Lu et al. (2021) | Price, vehicle relocation, mode choice | A bilevel nonlinear mathematical programming model, solved by genetic algorithm | Modelled using a binary logit model |
| Automated taxis | Hörl et al. (2021) | Price, waiting time, fleet size, demand | An agent-based simulation | Modelled using a multinomial logit model |
| E-hailing services integrated with transit service | Zhang and Nie (2021) | Mode choice, passenger demand, vehicle supply, pricing strategies for riders, wage for drivers, passenger waiting time | An aggregate equilibrium model, solved using a simple iterative fixed-point algorithm | Modelled using a multinomial logit model |
| Dial-a-ride problem | Dong et al. (2022) | Fleet size, accept/reject decisions, trip assignment, and vehicle routing | A mixed-integer programming model, solved by two customised heuristic algorithms | Modelled using logit-based chance constraints |
| Public transit planning integrated with ride-hailing | Wei et al. (2022) | Public transit schedule and operations, mode choices, ride-hailing operations, traffic congestion | A mixed-integer nonlinear programming model, solved by a bilevel decomposition algorithm | Modelled using a multinomial logit model |
| SAV management | Guo et al. (2022) | Vehicle operation, mode choice | A mixed-integer nonlinear programming model, reformulated and then solved by a solver | Modelled using a binary logit model |
| Competition between multiple AMoD systems | Wang et al. (2022) | Fleet sizes, assignment strategies, pricing strategies, mode choice | An agent-based simulation | Modelled using a multinomial logit model |
| SAV fleet sizing and management | Fan et al. (2023) | Fleet size, initial fleet distribution, mode choice, service quality, congestion, trip assignment, and vehicle operation | A mixed-integer nonlinear programming model, reformulated and then solved by a solver | Modelled using a binary logit model |
| Private vehicles, buses, and solo and pooled ride-hailing, with pooled vehicles allowed to use dedicated bus lanes | Fayed et al. (2023) | Solo–pool demand split, proportion of pooled vehicles using bus lanes, and pooling discounts | A static analytical macroscopic equilibrium model and system-optimum model | Modelled using binary logit and nested-logit model |
| Private vehicles, buses, and solo and pooled ride-hailing, with optional bus-lane access for pooled vehicles | Fayed et al. (2024) | Dynamic pooling fares and allocation of pooled vehicles between general-purpose and bus-lane networks | A dynamic macroscopic model with PI and model predictive control | Modelled using multinomial logit model |

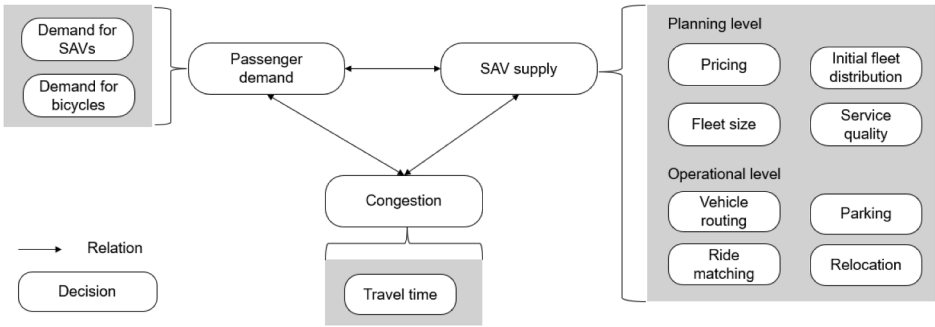



**Fig. 1.** Model variables and their interactions.

(d) Travellers are assumed to use only one mode of transportation; mode transfers are not considered.
(e) Ride-pooling services are excluded from consideration in this research.

### 3.2. Model setting

The problem is formulated as a mathematical model that optimises planning decisions by capturing the interactions among passenger demand, SAV supply, and traffic congestion. Fig. 1 illustrates the key variables in the model and their interactions. Some variables represent primary decision variables controlled by the operator (e.g. fleet size and pricing); others are auxiliary variables endogenously determined by the model constraints (e.g. congestion, represented through travel time).

We consider peak-hour traffic on a typical workday in an urban area by aggregating trips with identical travel information and socio-demographic information into groups. The travel data includes origin, destination, departure time, and latest arrival time. For socio-demographic information, we categorise travellers based on three income levels: high, middle, and low. Travellers within each group have the freedom to choose their mode of transportation (either SAV or bicycle). Due to the congestion effects, SAVs serving travellers in the same groups are allowed to choose different routes to avoid the congestion. The sets and parameters used to detail this trip information are the following.

| | |
|---|---|
| Sets associated with trips | |
| $C$ | Set of income classes, categorised into low income ($low$), middle income ($mid$), and high income ($high$). |
| $\mathcal{R}^c$ | Set indexing groups of trips associated with a specific income class $c \in C$. Trips within the same group $r \in \mathcal{R}^c$ share the same characteristics, including origin, destination, departure time, latest arrival time, and income level $c$. |
| $\mathcal{M}$ | Set of travel modes, consisting of automated vehicles ($AV$) and bicycles ($B$) as options. |
| Parameters associated with trips | |
| $o^r$ | Origin node for group of trips $c \in C, r \in \mathcal{R}^c$. |
| $d^r$ | Destination node for group of trips $c \in C, r \in \mathcal{R}^c$. |
| $a^r$ | Departure time for group of trips $c \in C, r \in \mathcal{R}^c$. |
| $b^r$ | Latest arrival time for group of trips $c \in C, r \in \mathcal{R}^c$. |
| $l^r_{\text{short}}$ | Shortest travel distance for group of trips $c \in C, r \in \mathcal{R}^c$, in kilometres. |
| $t^r_{\text{short}}$ | Shortest travel time assuming free-flow speed for group of trips $c \in C, r \in \mathcal{R}^c$, in time steps. |
| $n^r$ | Total number of trips for group $c \in C, r \in \mathcal{R}^c$. |

We employ a time-space network to address dynamic operational decision-making and the endogenous traffic congestion resulting from the large-scale deployment of SAVs. This network simultaneously represents spatial movements and temporal progression. Nodes in the time-space network correspond to physical nodes $i \in \mathcal{N}$ at discrete time instant $t \in \mathcal{T}$, where the length of each time step is denoted by $\Delta t$. Links represent possible transitions between these nodes, denoted as $(i_{t_1}, j_{t_2}) \in \mathcal{G}$, where road link $(i, j) \in \mathcal{L}, t_1, t_2 \in \mathcal{T}, t_1 < t_2, t^{\min}_{ij} \leq t_2 - t_1 \leq t^{\max}_{ij}$. The sets and parameters associated with the network are summarised below.

| | |
|---|---|
| Sets associated with the network setting | |
| $\mathcal{T}$ | Set of discrete time instants $\mathcal{T} = \{0, 1, 2, \dots, T\}$ in the operational period. |
| $\mathcal{N}$ | Set of physical nodes within the network. |
| $\mathcal{L}$ | Set of road links connecting the nodes in set $\mathcal{N}$. |
| $\mathcal{G}$ | Set of links in the time-space network. |
| $\mathcal{N}_P$ | Subset of nodes that allow parking for SAVs, where $\mathcal{N}_P \subseteq \mathcal{N}$. |
| Parameters associated with the network setting | |
| $\Delta t$ | Time step representing one time unit in the time-space network. |
| $l_{ij}$ | Length of road link $(i, j) \in \mathcal{L}$, in kilometres. |
| $Q_{ij}$ | Capacity of road link $(i, j) \in \mathcal{L}$ in number of vehicles per time step. |
| $t_{ij}^{\max}$ | Maximum travel time by car on road link $(i, j) \in \mathcal{L}$, in time steps. |
| $t_{ij}^{\min}$ | Minimum travel time by car on road link $(i, j) \in \mathcal{L}$, in time steps |
| $C_{i_{t_1} j_{t_2}}$ | Spatial capacity of road link $(i, j) \in \mathcal{L}$ in number of vehicles that fit on the road link from time instant $t_1$ to $t_2$, where $(i_{t_1}, j_{t_2}) \in \mathcal{G}$. |

### 3.3. Passenger demand modelling

We employ a discrete choice model to express travel mode preferences among travellers. Utility values for the two modes considered in this paper, SAVs and bicycles, are computed. We use a binary logit model to estimate the probability of using each mode for different groups of trips.

**Binary logit model:** This section presents the utility functions and the corresponding binary logit model equations.

Income heterogeneity has been widely demonstrated in the literature to influence various aspects of travel behaviour, including individuals' sensitivity to monetary costs, their value of travel time, perceptions of travel time components such as walking and waiting, and their comfort preferences (e.g., Kolarova et al., 2019). These multidimensional effects highlight the importance of explicitly representing income differences in demand modelling to capture realistic behavioural heterogeneity.

The utility of using mode $m \in \mathcal{M}$ for travellers in group $c \in C, r \in \mathcal{R}^c$ is expressed as $U_m^r = V_m^r + \epsilon_m^r$. Here, $V_m^r$ represents the deterministic component of utility, which is a function of the observed attributes of the alternatives and the characteristics of the individual. The term $\epsilon_m^r$ captures the random component of utility, reflecting unobserved influences on choice behaviour (Ben-Akiva et al., 1985). Assuming that $\epsilon_m^r$ is independently and identically distributed according to a Type I extreme value (Gumbel) distribution (Train, 2009), we can obtain the probability $P_{AV}^r$ of using SAVs for groups $c \in C, r \in \mathcal{R}^c$ using a binary logit model, as shown in Eq. (1). To determine the probability of using SAVs, we need to know the deterministic utility $V_{AV}^r$ of SAVs and $V_B^r$ of bicycles, which can be expressed as Eqs. (2) and (3), respectively.

$$P_{AV}^r = \frac{e^{V_{AV}^r}}{e^{V_{AV}^r} + e^{V_B^r}}, \qquad \forall r \in \mathcal{R}^c, c \in C \tag{1}$$

$$V_{AV}^r = -\beta_0^c (c_{AV}^r + VOT_{AV}^c T_{AV}^r) - \beta_1 (1 - \alpha), \qquad \forall c \in C, r \in \mathcal{R}^c \tag{2}$$

$$V_B^r = -\beta_0^c (c_B^r + VOT_B^c T_B^r), \qquad \forall c \in C, r \in \mathcal{R}^c \tag{3}$$

The deterministic utility of using SAVs, denoted by the continuous variable $V_{AV}^r$, consists of three components as given in Eq. (2): the fare for the services, travel time-related costs, and traveller satisfaction concerning the trip rejection rate. To capture the sensitivity of travellers from different income classes to changes in costs (measured in euros), we introduce a class-specific parameter, $\beta_0^c$. As indicated in Kolarova et al. (2019), individuals with high income tend to perceive costs less negatively than those with middle or lower incomes, with middle-income individuals perceiving costs less negatively than those with low incomes. The cost for group $c \in C, r \in \mathcal{R}^c$ is determined as the monetary cost of using SAV services $c_{AV}^r$ and the travel time-related cost $VOT_{AV}^c T_{AV}^r$. Here, $VOT_{AV}^c$ is a parameter representing the income class-based value of travel time (VOTT) for using AVs. The estimated VOTT for individuals with high income is higher than that of those in middle- or low-income classes, and the VOTT for the middle-income group is higher than that of the low-income group (Kolarova et al., 2019). $T_{AV}^r$ are integer variables representing the longest travel time of using AVs by group $r \in \mathcal{R}^c, c \in C$. Trips can be rejected at a rate of $1 - \alpha$, where $\alpha$ is a continuous variable in the range [0,1]. The sensitivity of travellers to this is described by the parameter $\beta_1$. Rejecting trips negatively impacts people's willingness to use SAV services.

The deterministic term $V_B^r$ of the utility for using a bicycle for group $c \in C, r \in \mathcal{R}^c$ comprises two parts: the monetary cost parameter $c_B^r$ which reflects the bicycle's depreciation cost, and the travel time-related costs $VOT_B^c T_B^r$. Since bicycle travel times $T_B^r$ are less affected by congestion due to the flexibility of bicycles, it is treated as a parameter rather than a variable. Therefore, the utility $V_B^r$ for using a bicycle can be represented by Eq. (3). Travellers' waiting time for the SAV services is not explicitly included in the utility function, as this study focuses on strategic planning under the assumption of complete knowledge of trip demand. Our objective is to determine the minimum number of vehicles needed to ensure on-time departures for travellers choosing between SAVs and bicycles. Given this limited mode choice, we expect SAV services to offer relatively short waiting times. Explicitly modelling waiting time would increase the model's complexity and make it hard to solve. While this simplification may limit the model's applicability in on-demand scenarios where passengers request immediate service with little advance notice, it enables a more tractable planning approach.

**Demand calculation:** This section presents the constraints used to calculate demand.

Knowing the probability of using SAVs, we can determine the SAV demand for each group $r \in \mathcal{R}^c, c \in C$. This demand is represented by integer variable $D^r_{AV}$ and is computed using Constraints (4). This variable is computed as the product of the total number of trips $n^r$ in group $c \in C, r \in \mathcal{R}^c$ and the continuous probability variable $P^r_{AV}$ of choosing SAVs. The addition and subtraction of 0.5 on either side of Constraints (4) ensures that the demand will take the integer value closest to the value of $n^r P^r_{AV}$.

$$n^r P^r_{AV} - 0.5 < D^r_{AV} \leq n^r P^r_{AV} + 0.5, \qquad \forall c \in C, r \in \mathcal{R}^c. \tag{4}$$

### 3.4. SAV service planning and operation modelling

In this subsection, we first present the modelling of pricing strategies in Section 3.4.1. We then introduce the mathematical models for fleet management (fleet sizing, initial fleet distribution and service quality) and operations (trip assignment and vehicle routing) in Section 3.4.2.

#### 3.4.1. Pricing strategy

From the perspective of the SAV service provider, we aim to explore three distinct pricing strategies: (1) charging travellers an initial base fare followed by a distance-related fare; (2) charging travellers solely based on their travel distance; (3) implementing a pricing strategy that varies according to travellers' income class level, which is an income class-based, distance-related pricing scheme. The detailed explanation and equations for these three pricing strategies are presented below.

**Pricing strategy 1. Base fare + distance-based fare**

This pricing strategy is among the most commonly employed. The base fare, denoted by continuous variable $p^0$, a fixed amount charged at the ride's outset, ensures drivers earn a minimum amount per trip from the moment they accept a ride. This is crucial for short trips, where time and distance fares may not sufficiently compensate the driver. Additionally, it may deter the use of SAVs for short distances. Introducing a base fare, as discussed in this paper, can improve bicycle use, offering a more eco-friendly transportation alternative. Following the base fare $p^0$, a distance-based fare, denoted by continuous variable $p$, is introduced based on the shortest travel distance $l^r_{\text{short}}$, a parameter pre-calculated using the Dijkstra algorithm (Dijkstra, 1959). The out-of-pocket money $c^r_{AV}$ for travellers in group $r \in \mathcal{R}^c$ belonging to class $c \in C$ is obtained following Eq. (5).

$$c^r_{AV} = p^0 + l^r_{\text{short}} p, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{5}$$

**Pricing strategy 2. Distance-based fare**

Distance-based pricing is commonly used in some ride-sharing platforms, delivery services, and some traditional taxi services. It offers a straightforward price structure, which is favoured in markets where consumers prefer predictable pricing structures. The out-of-pocket money paid by travellers will be obtained using a distance-based fare $p$ times the shortest travel distance $l^r_{\text{short}}$, as shown in Eq. (6).

$$c^r_{AV} = l^r_{\text{short}} p, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{6}$$

**Pricing strategy 3. Income class-based fare**

Income class-based pricing adjusts fares based on the income level of travellers. For simplicity, we focus solely on income class-based distance-based pricing strategies. The out-of-pocket money for travellers in group $c \in C, r \in \mathcal{R}^c$ is expressed in Eq. (7), where $p^c$ is a continuous variable representing distance-based fare for income class $c \in C$.

$$c^r_{AV} = l^r_{\text{short}} p^c, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{7}$$

#### 3.4.2. Fleet management and operation

This section outlines the constraints that describe detailed SAV operations, including trip assignment, routing, and parking. Based on these operations and the endogenous SAV demand, service providers can make the most rational fleet management decisions. These include determining fleet size, setting the initial distribution of the fleet across parking depots, and ensuring service quality. The constraints are detailed as follows.

**Service rate calculation:** This section introduces the constraints that define demand and service rate calculation.

SAV operators can reject those requests that bring no profits to the company. Under this situation, the number of served trips $S^r$, an integer variable for group $c \in C, r \in \mathcal{R}^c$, can be less than or equal to the group's demand $D^r_{AV}$, which is also an integer variable. This is ensured by Constraints (8). Constraint (9) determines the endogenously determined continous variable service rate $\alpha$ of the SAV service as the total number of served trips divided by the total demand for SAV. This is a non-linear constraint because of the product of continuous variable $\alpha$ and integer variables $D^r_{AV}$.

$$S^r \leq D^r_{AV}, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{8}$$

$$\alpha \sum_{c \in C, r \in \mathcal{R}^c} D^r_{AV} = \sum_{c \in C, r \in \mathcal{R}^c} S^r \tag{9}$$

**Trip assignment and vehicle routing constraints:** This section introduces the constraints that describe passenger movements, trip assignment, and vehicle movements within the network, ensuring that vehicles follow feasible routes and serve all assigned trips.

To describe the passenger flow movement on the network, we define continuous variables $\hat{F}^r_{i_{t_1} j_{t_2}}$ which represent the passenger flow in group $c \in C, r \in \mathcal{R}^c$ served by an SAV on road link $(i, j) \in \mathcal{L}$, from time instant $t_1 \in \mathcal{T}$ to $t_2 \in \mathcal{T}$. This variable is defined only

for $(i_{t_1}, j_{t_2}) \in \mathcal{G}$, with $a^r \le t_1 < t_2 \le b^r$, where $a^r$ and $b^r$ are the departure time and the latest arrival time for passenges in group $c \in C, r \in \mathcal{R}^c$. If $t_1 = a^r$, then the starting node $i$ corresponds to the origin $o^r$. Constraints (10) ensure that passengers in group $c \in C, r \in \mathcal{R}^c$ are picked up by the SAVs at the origin node $o^r$ at departure time $a^r$. After departure, SAVs are allowed to choose different paths to avoid traffic congestion, even though they serve passengers with the same income class and in the same group.

To capture the different arrival times of the SAVs, Constraints (11) and (12) are defined to ensure that the number of served trips $S^r$ in group $c \in C, r \in \mathcal{R}^c$ is equal to the number of trips arriving at destination $d^r$ at time $t$ with $t \ge a^r + t^r_{\text{short}}$ and $t \le b^r$, denoted by variable $E^r_t$. Constraints (13) and (14) ensure that the passengers are left at the destination node $d^r$ upon the SAV's first arrival, and there will be no return trips to the origin node $o^r$ by the SAVs after they depart.

The vehicle flow movement is represented by continuous variables $F_{i_{t_1} j_{t_2}}$, which denote the vehicle flow in road link $(i, j)$ from time instant $t_1$ to $t_2$, where $t_1 < t_2$, $t^{\min}_{ij} \le t_2 - t_1 \le t^{\max}_{ij}$, and $(i_{t_1}, j_{t_2}) \in \mathcal{G}$. Constraints (15) make sure that the vehicle flow circulation $F_{i_{t_1} j_{t_2}}$ on the time-space network must be greater than or equal to the passenger flow $\hat{F}^r_{i_{t_1} j_{t_2}}$. The vehicle flows describe not only the delivery process of a passenger but also the empty relocation process.

$$S^r = \sum_{j_t | \left(o^r_{a^r}, j_t\right) \in \mathcal{G}} \hat{F}^r_{o^r_{a^r} j_t}, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{10}$$

$$S^r = \sum_{t \in \mathcal{T} | a^r + t^r_{\text{short}} \le t \le b^r} E^r_t, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{11}$$

$$E^r_t = \sum_{i_{t_1} | \left(i_{t_1}, d^r_t\right) \in \mathcal{G}} \hat{F}^r_{i_{t_1} d^r_t}, \qquad \forall c \in C, r \in \mathcal{R}^c, a^r + t^r_{\text{short}} \le t \le b^r \tag{12}$$

$$\sum_{j_{t_1} | \left(d^r_t, j_{t_1}\right) \in \mathcal{G}} \hat{F}^r_{d^r_t j_{t_1}} = 0, \qquad \forall c \in C, r \in \mathcal{R}^c, a^r + t^r_{\text{short}} \le t \le b^r \tag{13}$$

$$\sum_{i_{t_1} | \left(i_{t_1}, o^r_t\right) \in \mathcal{G}} \hat{F}^r_{i_{t_1} o^r_t} = 0, \qquad \forall r \in \mathcal{R}, c \in C, a^r \le t \le b^r \tag{14}$$

$$\sum_{r \in \mathcal{R}, c \in C} \hat{F}^r_{i_{t_1} j_{t_2}} \le F_{i_{t_1} j_{t_2}}, \qquad \forall \left(i_{t_1}, j_{t_2}\right) \in \mathcal{G} \tag{15}$$

**Flow conservation constraints:** This section presents the constraints that ensure the conservation of passenger and vehicle flows in the network.

Constraints (16)–(18) describe the passenger flow conservation rule at node $i \in \mathcal{N}$ in the network and the vehicle flow conservation at both normal and parking nodes. For parking nodes, $W_{i_t}$ denotes total number of vehicles parking at node $i \in \mathcal{N}_P$ from time instant $t$ to $t+1$, with $t \in \mathcal{T} \setminus \{T\}$.

$$\sum_{j_{t_1} | \left(j_{t_1}, i_t\right) \in \mathcal{G}} \hat{F}^r_{j_{t_1} i_t} = \sum_{j_{t_2} | \left(i_t, j_{t_2}\right) \in \mathcal{G}} \hat{F}^r_{i_t j_{t_2}}, \qquad \forall c \in C, r \in \mathcal{R}^c, a^r < t < b^r, i \in \mathcal{N}, i \ne o^r, i \ne d^r \tag{16}$$

$$\sum_{j_{t_1} | \left(j_{t_1}, i_t\right) \in \mathcal{G}} F_{j_{t_1} i_t} = \sum_{j_{t_2} | \left(i_t, j_{t_2}\right) \in \mathcal{G}} F_{i_t j_{t_2}}, \qquad \forall i \in \mathcal{N} \setminus \mathcal{N}_P, 0 < t < \mathcal{T} \tag{17}$$

$$\sum_{j_{t_1} | \left(j_{t_1}, i_t\right) \in \mathcal{G}} F_{j_{t_1} i_t} + W_{i_{t-1}} = \sum_{j_{t_2} | \left(i_t, j_{t_2}\right) \in \mathcal{G}} F_{i_t j_{t_2}} + W_{i_t}, \qquad \forall i \in \mathcal{N}_P, 0 < t < \mathcal{T} \tag{18}$$

**Fleet sizing and distribution constraints:** This section explains the constraints that determine the total fleet size and the distribution of vehicles at the beginning of the optimisation period.

Constraints (19) specify the initial fleet distribution $O_i$, an integer variable, across parking nodes $i \in \mathcal{N}_P$ at the beginning of the studied period. When the optimisation period starts, these SAVs either set out from their parking locations to pick up passengers (indicated by integer variable $F_{i_0 j_t}$) or stay at the parking node (indicated by integer variable $W_{i_0}$), ready to receive orders from the SAV operator. The total SAV fleet size $O$, also an integer variable, is expressed as the sum of the initial fleet distribution $O_i$ at each parking node $i \in \mathcal{N}_P$, which is ensured by Constraint (20).

$$\sum_{j_t | (i_0, j_t) \in \mathcal{G}} F_{i_0 j_t} + W_{i_0} = O_i, \qquad \forall i \in \mathcal{N}_P \tag{19}$$

$$\sum_{i \in \mathcal{N}_P} O_i = O \tag{20}$$

**Longest travel time determination:** This section presents the constraints used to determine the longest travel time for each group of trips.

When evaluating the utility of using SAVs for group $r \in \mathcal{R}^c$, $c \in C$, we consider the longest travel time $T^r_{AV}$ among all travellers in the group. This is because congestion may cause travellers in the same group to experience different travel times, and the longest travel time reflects the delay that affects the group most. Constraints (21)–(25) are introduced to obtain this value. The longest travel time should be greater than or equal to the shortest free-flow travel time $t^r_{\text{short}}$, and less than or equal to the maximum possible travel

time $b^r - a^r$, as ensured by Constraints (21). To capture the exact arrival time for each trip, we introduce binary variables $A_t^r$, which takes value 1 if there are trips in group $c \in C, r \in \mathcal{R}^c$ that arrive at time instant $t \in \mathcal{T}, a^r + t^r_{\text{short}} \le t \le b^r$, and 0 otherwise. Constraints (22) define the upper and lower bounds for these binary variables. Constraints (23) ensure that the longest travel time variable $T^r_{AV}$ is greater than or equal to the travel times of all trips within a group. To guarantee that $T^r_{AV}$ takes only one value that is the actual longest travel time, we introduce binary variables $Z_t^r$, where $Z_t^r$ take value 1 when the latest arrival time for group $c \in C, r \in \mathcal{R}^c$ occurs at time instant $t \in \mathcal{T}, a^r + t^r_{\text{short}} \le t \le b^r$. Constraints (24) and (25) ensure that only one time instant is selected as the longest travel time and an upper bound is imposed on $T^r_{AV}$. In Constraints (25), $M^r$ is a sufficiently large number. To tighten the model, $M^r$ can be set to $b^r - a^r$ for $c \in C, r \in \mathcal{R}^c$.

$$t^r_{\text{short}} \le T^r_{AV} \le b^r - a^r, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{21}$$

$$\frac{E_t^r}{n^r} \le A_t^r \le E_t^r, \qquad \forall c \in C, r \in \mathcal{R}^c, a^r + t^r_{\text{short}} \le t \le b^r \tag{22}$$

$$T^r_{AV} \ge A_t^r (t - a^r), \qquad \forall c \in C, r \in \mathcal{R}^c, a^r + t^r_{\text{short}} \le t \le b^r \tag{23}$$

$$\sum_{t | a^r + t^r_{\text{short}} \le t \le b^r} Z_t^r = 1, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{24}$$

$$T^r_{AV} \le A_t^r (t - a^r) + M^r (1 - Z_t^r), \qquad \forall c \in C, r \in \mathcal{R}^c, a^r + t^r_{\text{short}} \le t \le b^r \tag{25}$$

### 3.5. Traffic congestion modelling

The Bureau of Public Roads (BPR) function is one of the most widely used volume-delay functions in transportation planning (Seshadri et al., 2022; Ke and Qian, 2023). It provides a method for estimating link travel times as traffic volume increases. In our model, we use a variation of the BPR function by introducing time-dependent link capacity, which is the so-called spatial link capacity introduced in Van Essen and Correia (2019). Instead of including the traditional non-linear BPR function (Dafermos and Sparrow, 1969) directly into the model to determine the travel time, they define spatial link capacity which is the maximum number of vehicles that can pass road link $(i,j) \in \mathcal{L}$ from time instant $t_1 \in \mathcal{T}$ to time instant $t_2 \in \mathcal{T}$ with $t^{\min}_{ij} \le t_2 - t_1 \le t^{\max}_{ij}$, denoted by $C_{i_{t_1} j_{t_2}} = (t_2 - t_1) Q_{ij} \left( \frac{1}{a} \left( \frac{t_2 - t_1}{t^{\min}_{ij}} - 1 \right) \right)^{\frac{1}{b}}$. Here, $Q_{ij}$ represents the practical capacity of road link $(i,j) \in \mathcal{L}$ measured in number of vehicles that can flow through the link. $a$ and $b$ are parameters.

Instead of directly determining the travel time, we select an option from multiple link-time-capacity combinations, as described by Constraints (26) and (27). Binary variables $X_{i_{t_1} j_{t_2}}$ are introduced, which indicate whether a vehicle travels on road link $(i,j)$ from time instant $t_1$ to $t_2$, with $(i_{t_1}, j_{t_2}) \in \mathcal{G}$. Constraints (26) ensure that at most one travel time option is selected for link $(i,j) \in \mathcal{L}$ at any given time instant $t \in \mathcal{T}$. Constraints (27) ensure that the corresponding vehicle flow $F_{i_{t_1} j_{t_2}}$ does not exceed the spatial capacity of the road $C_{i_{t_1} j_{t_2}}$ when travel occurs on road link $(i,j)$ from time instant $t_1$ to $t_2$.

Please note that our model does not consider queueing, signals at intersections, and microscopic traffic dynamics for simplification purposes. Incorporating these factors into the link travel time estimation will significantly increase data requirements and computational complexity, making the optimisation problem much more difficult to solve. Despite this simplification, we do believe that for the strategic purposes of the model, the BPR function is acceptable.

**Capacity constraints**

$$\sum_{t_1 | \left(i_t, j_{t_1}\right) \in \mathcal{G}} X_{i_t j_{t_1}} \le 1, \qquad \forall (i,j) \in \mathcal{L}, t \in \mathcal{T}, t < T \tag{26}$$

$$F_{i_{t_1} j_{t_2}} \le \left\lfloor C_{i_{t_1} j_{t_2}} \right\rfloor X_{i_{t_1} j_{t_2}}, \qquad \forall \left(i_{t_1}, j_{t_2}\right) \in \mathcal{G} \tag{27}$$

Constraints (28) describe the first-in-first-out (FIFO) rule, meaning that the vehicles that enter the road link first will leave the road link first. $M_{ijt_1t_2}$ is a sufficiently large number. To tighten the model, $M_{ijt_1t_2}$ can be set to $M_{ijt_1t_2} = t_1 + t^{\max}_{ij} - t_2$, for $(i,j) \in \mathcal{L}, t_1 < t_2 \le t_1 + t^{\max}_{ij} - t^{\min}_{ij}$. This ensures that when $\sum_{t \in \mathcal{T}} X_{i_{t_2} j_t}$ equals 0, the maximum value of the left-hand side of Constraints (28), $t_1 + t^{\max}_{ij}$, will still be less than or equal to the right-hand side of Constraints (28).

**FIFO constraints**

$$t_1 + \sum_{t \in \mathcal{T}} X_{i_{t_1} j_t} (t - t_1) \le t_2 + \sum_{t \in \mathcal{T}} X_{i_{t_2} j_t} (t - t_2) + M_{ijt_1t_2} \left( 1 - \sum_{t \in \mathcal{T}} X_{i_{t_2} j_t} \right), \qquad \forall (i,j) \in \mathcal{L}, \; t_1 < t_2 \le t_1 + t^{\max}_{ij} - t^{\min}_{ij} \tag{28}$$

### 3.6. Objective function

To maximise the SAV operator's profit, we establish an objective function (see Eq. (29)) that accounts for the service's total income and its operational expenses. The total income is represented by the sum of the fares paid by all travellers, which is the first term of the objective function. It is crucial to highlight that this sum includes non-linear terms, which are the multiplications of continuous variables $c^r_{AV}$ and integer variables $S^r$. Costs are divided into: (1) fleet depreciation, which is the purchase cost per vehicle divided by its lifespan given by $c_f$, multiplied by the fleet size, and is represented by the second term in Eq. (29); (2) operational costs, including fuel, maintenance, and insurance, determined by multiplying the total distance travelled by all SAVs by the cost per kilometre $c_o$, as indicated by the third term in Eq. (29); and (3) penalties for late drop-offs $c_d$, charged per time unit for delays beyond the shortest anticipated travel time, shown by the final term in Eq. (29). Even though delay is indirectly captured in the utility function (Eq. (2)), we include an explicit delay penalty in the objective function to ensure that the service provider is incentivised to minimise excessive delays for all users, rather than only prioritising profitability. Without this term, the model could trade off higher delays for certain user groups in exchange for increased revenue and improved fleet utilisation, potentially leading to inequitable outcomes.

$$\max \sum_{c\in C, r\in\mathcal{R}^c} c^r_{AV} S^r - c_f \cdot O - c_o \left( \sum_{\left(i_{t_1}, j_{t_2}\right)\in\mathcal{G}} l_{ij} F_{i_{t_1} j_{t_2}} \right) - c_d \sum_{c\in C, r\in\mathcal{R}^c} \left( \sum_{t\in\mathcal{T}} t E^r_t - (a^r + t^r_{\text{short}}) S^r \right) \tag{29}$$

The proposed models for the three pricing strategies are summarised below and are denoted as **[M0-1]**, **[M0-2]**, and **[M0-3]**, respectively.

*Pricing strategy 1: base fare + distance-based fare.* **[M0-1]** Objective function (29), subject to Constraints (1)–(5) and (8)–(28).

*Pricing strategy 2: distance-based fare.* **[M0-2]** Objective function (29), subject to Constraints (1)–(4), (6) and (8)–(28).

*Pricing strategy 3: income class-based fare.* **[M0-3]** Objective function (29), subject to Constraints (1)–(4) and (7)–(28).

## 4. Solution algorithms

The model presented in Section 3 is a MINLP model, which is challenging to solve. This nonlinearity arises from three aspects: the nonlinear binary logit model described in Eq. (1); the nonlinear constraint (9) for calculating the service rate, which involves the multiplication of a continuous variable with integer variables; and the nonlinear objective function (29) that includes products of continuous and integer variables.

This paper proposes three distinct solution strategies to address the aforementioned nonlinearities: (1) reformulating the model into a Mixed-Integer Linear Programming (MILP) model by applying linearisation techniques, introduced in Section 4.1; (2) combining Particle Swarm Optimisation (PSO) with a reformulated MILP in an iterative framework; and (3) combining Bayesian Optimisation (BO) with a reformulated MILP.

### 4.1. Reformulated MILP model (M1)

Fan et al. (2023) propose a solution method to tackle the non-linearities brought by the binary logit model and the non-linear constraints. They transform the binary logit model into equations containing logarithmic terms and approximate these terms using an outer-inner approximation technique. Furthermore, they employ binary variables and the big-M method to reformulate the products of a continuous variable and integer variables. In this paper, we employ these techniques to linearise Eq. (1) and Constraints (9), as outlined by Constraints (A.29)–(A.51) in Appendix A. We further linearise the nonlinear terms in the objective function, which involves the product of continuous and integer variables. The continuous variables represent pricing, while the integer variables represent the number of trips served. We now describe the linearisation of the objective function for the three proposed pricing strategies, followed by the presentation of the overall mathematical formulation for this method. This method is referred to as M1 throughout the paper.

#### 4.1.1. Objective function linearisation with different pricing strategies

The nonlinear objective function is shown in Eq. (29) with three different pricing strategies described by Eqs. (5), (6), and (7). To linearise the first term in the objective function, we firstly introduce binary variables $\overline{S}^r_h$ to discretise the integer variable $S^r$ for each group $c \in C, r \in \mathcal{R}^c$, as shown in Constraints (30). Here, parameter $H^r = \lfloor \log_2(n^r) \rfloor$ ensures that Constraints (30) still hold when all the trips $n^r$ in group $c \in C, r \in \mathcal{R}^c$ would use SAVs.

$$S^r = \sum_{h=0}^{H^r} 2^h \overline{S}^r_h, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{30}$$

Then, the non-linear term $\sum_{r\in\mathcal{R}^c, c\in C} c^r_{AV} S^r$ can be reformulated as $\sum_{r\in\mathcal{R}^c, c\in C} \sum_{h=0}^{H^r} 2^h c^r_{AV} \overline{S}^r_h$. We additionally introduce continuous variables $\overline{Y}^r_h$ to replace the still non-linear term $c^r_{AV} \overline{S}^r_h$. Subsequently, we introduce the following constraints. Here, $p^r_{\max}$ represents the upper-bound value of the variable $c^r_{AV}$ with $r \in \mathcal{R}^c, c \in C$.

$$\overline{Y}^r_h \le p^r_{\max} \overline{S}^r_h, \qquad \forall c \in C, r \in \mathcal{R}^c, h \in \{0, 1, \dots, H^r\} \tag{31}$$

$$\overline{Y}^r_h \geq c^r_{AV} - p^r_{\max}(1 - \overline{S}^r_h), \qquad \forall c \in C, r \in \mathcal{R}^c, h \in \{0, 1, \dots, H^r\} \tag{32}$$

$$\overline{Y}^r_h \geq 0, \qquad \forall c \in C, r \in \mathcal{R}^c, h \in \{0, 1, \dots, H^r\} \tag{33}$$

$$\overline{Y}^r_h \leq c^r_{AV}, \qquad \forall c \in C, r \in \mathcal{R}^c, h \in \{0, 1, \dots, H^r\}. \tag{34}$$

The objective function is then updated as follows:

$$\max \sum_{r \in \mathcal{R}, c \in C} \sum_{h=0}^{H^r} 2^h \overline{Y}^r_h - c_f \cdot O - c_o \left( \sum_{\left(i_{t_1}, j_{t_2}\right) \in \mathcal{G}} l_{ij} F_{i_{t_1} j_{t_2}} \right) - c_d \sum_{c \in C, r \in \mathcal{R}^c} \left( \sum_{t \in \mathcal{T}} t E^r_t - (a^r + t^r_{\text{short}}) S^r \right). \tag{35}$$

We now describe the process for appropriately determining the value of $p^r_{\max}$, which is the upper bound of the variable $c^r_{AV}$. The variable $c^r_{AV}$ denotes the monetary costs for using SAV service for travellers in group $c \in C, r \in \mathcal{R}^c$. The monetary costs associated with using SAVs must be kept within a reasonable range to ensure the service remains attractive to users. If these costs are set too high, the SAV service may become unaffordable, leading all users to opt for bicycles instead. Our objective, therefore, is to identify the maximum value of $c^r_{AV}$ that still sustains at least one trip of demand for SAVs across at least one group of trips. In other words, the upper bound on the price should be high enough to optimise revenue without causing a complete loss of demand for SAV services within the entire system.

According to Eq. (4), the demand $D^r_{AV}$ becomes zero if $n^r P^r_{AV} < 0.5$. To maintain a non-zero demand for SAVs, $P^r_{AV}$ must be at least $\frac{1}{2n^r}$. To ensure a non-zero demand for at least one group $c \in C, r \in \mathcal{R}^c$, there must be at least one group that satisfies Constraints (36). In these constraints, the left-hand side represents the probability of using SAVs for group $r$, which is derived from Eqs. (1), (2), and (3).

$$\frac{e^{-\beta^c_0 (c^r_{AV} + VOT^c_{AV} T^r_{AV}) - \beta_1 (1-\alpha)}}{e^{-\beta^c_0 (c^r_{AV} + VOT^c_{AV} T^r_{AV}) - \beta_1 (1-\alpha)} + e^{V^r_B}} \geq \frac{1}{2n^r}, \qquad \exists c \in C, r \in \mathcal{R}^c \tag{36}$$

By rewriting these constraints, we obtain the following:

$$c^r_{AV} \leq \frac{-\beta_1(1-\alpha) - V^r_B + \ln(2n^r - 1)}{\beta^c_0} - VOT^c_{AV} T^r_{AV}, \qquad \exists c \in C, r \in \mathcal{R}^c \tag{37}$$

The right-hand side of (37) is maximised when $\alpha = 1$ and $T^r_{AV}$ equals the minimum travel time $t^r_{\text{short}}$ at free-flow speed for group $c \in C, r \in \mathcal{R}^c$. Thus, instead of determining a price such that Constraint (37) holds for at least one group $r$, we take the maximum of the right-hand side over all groups and require $c^r_{AV}$ to be less than or equal to this value for all $r$.

$$c^r_{AV} \leq \max_{c \in C, r \in \mathcal{R}^c} \left( \frac{-V^r_B + \ln(2n^r - 1)}{\beta^c_0} - VOT^c_{AV} t^r_{\text{short}} \right), \qquad \forall c \in C, r \in \mathcal{R}^c \tag{38}$$

Thus, $p^r_{\max}$, which describes the upper bound of the variable $c^r_{AV}$, can take the value of the right-hand side of Constraints (38). Noted that for each $c \in C, r \in \mathcal{R}^c$, $p^r_{\max}$ has the same value, allowing us to simplify $p^r_{\max}$ to $p_{\max}$. The value of $p_{\max}$ is provided in Eq. (39). Consequently, Constraints (31) and (32) are updated to Constraints (40) and (41), respectively.

$$p_{\max} = \max_{c \in C, r \in \mathcal{R}^c} \left( \frac{-V^r_B + \ln(2n^r - 1)}{\beta^c_0} - VOT^c_{AV} t^r_{\text{short}} \right) \tag{39}$$

$$\overline{Y}^r_h \leq p_{\max} \overline{S}^r_h, \qquad \forall c \in C, r \in \mathcal{R}^c, h \in \{0, 1, \dots, H^r\} \tag{40}$$

$$\overline{Y}^r_h \geq c^r_{AV} - p_{\max}(1 - \overline{S}^r_h), \qquad \forall c \in C, r \in \mathcal{R}^c, h \in \{0, 1, \dots, H^r\} \tag{41}$$

The overall mathematical formulations for the three proposed pricing strategies are summarised as follows:

*Pricing strategy 1: base fare + distance-based fare.* **[M1-1]** Objective function (35) with Constraints (2)–(5), (8)–(28), (30), (33), (34), (38)–(41), the linearisation-related constraints (A.29)–(A.51), and the non-negativity constraints (A.52), (A.53), (A.59)–(A.78) in Appendix A.

*Pricing strategy 2: distance-based fare.* **[M1-2]** Objective function (35), subject to Constraints (2)–(4), (6), (8)–(28), (30), (33), (34), (38)–(41), the linearisation-related constraints (A.29)–(A.51), and the non-negativity constraints (A.53), (A.59)–(A.78) in Appendix A.

*Pricing strategy 3: income class-based fare.* **[M1-3]** Objective function (35), subject to Constraints (2)–(4), (7), (8)–(28), (30), (33), (34), (38)–(41), the linearisation-related constraints (A.29)–(A.51), and the non-negativity constraints (A.54)–(A.78) in Appendix A.

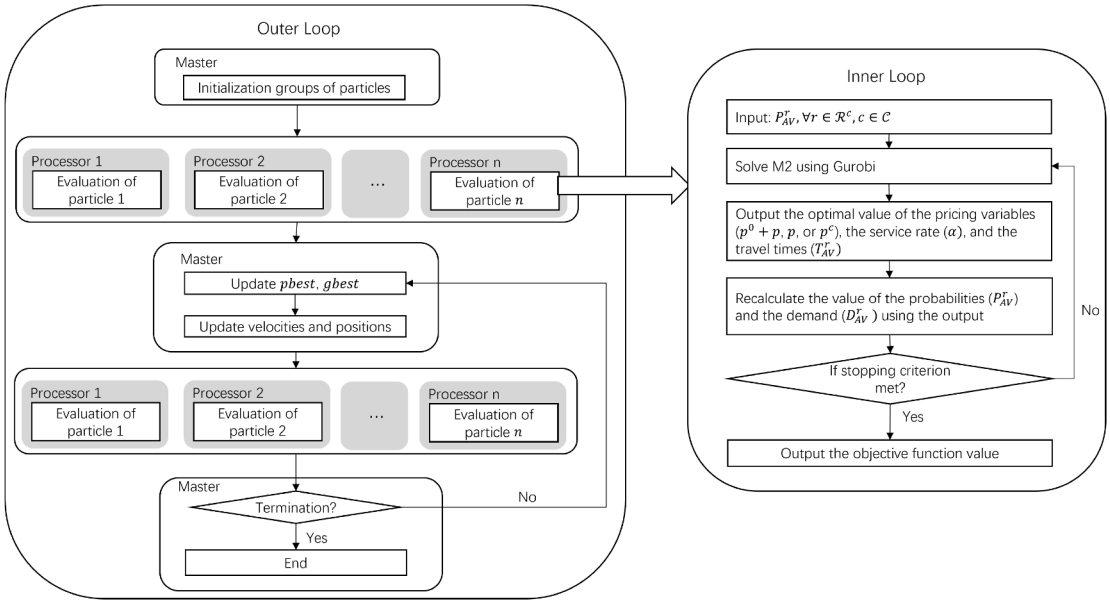



**Fig. 2.** Flowchart of the particle swarm optimisation framework embedded with an iterative process of solving a reformulated MILP model (M2).

### *4.2. Parallel particle swarm optimisation embedded with an iterative process of solving a refomulated MILP model (M2)*

Linearising the binary logit model with the outer-inner approximation method requires a significant number of binary variables and constraints. Specifically, the number of these variables and constraints expands with the growth in the number of groups, thereby escalating the computational load. In this section, we introduce a hybrid solution approach that integrates Particle Swarm Optimisation (PSO) with a reformulated MILP model, referred to as M2. In M2, the probability $P^r_{AV}$ of utilising SAV services for each group $c \in C, r \in \mathcal{R}^c$ is not treated as a variable, but as a fixed parameter provided as input. Solving M2 to optimality yields the optimal objective function value under these given probabilities. In PSO, we define the probabilities $P^r_{AV}$ for all groups $c \in C, r \in \mathcal{R}^c$ as a candidate solution, dubbed a particle. The primary objective of the PSO is to move particles around in the search space to find the best objective function value of M2.

We first present the framework of the proposed algorithm in Fig. 2. The detailed formulation of the reformulated MILP model (M2) is introduced in Section 4.2.1. Following this, Section 4.2.2 details the tailored PSO procedure specifically designed for our problem. The concepts and terms mentioned in Fig. 2, such as particles, *pbest*, *gbest*, velocities, and positions, are also described in Section 4.2.2.

#### *4.2.1. Inner loop: iterative process of solving a reformulated MILP model (M2)*

In the reformulated MILP model (M2), the pricing variables, fleet sizing variable, initial fleet distribution variables, and service rate variable are endogenous variables, while the probabilities of using SAVs for each group of trips are exogenous variables. Treating probabilities as exogenous variables enables us to transform the non-linear binary logit model into linear constraints within the model, thus facilitating an easier solution process. First, rewriting Eq. (1) and applying a logarithm gives us the following equations:

$$V^r_{AV} = \ln P^r_{AV} - \ln(1 - P^r_{AV}) + V^r_B, \qquad \forall c \in C, r \in \mathcal{R}^c. \tag{42}$$

Combining Eq. (42) with Eq. (2) establishes a link between the probabilities $P^r_{AV}$ of using SAV, the price variables $p^0, p, p^c$, travel time $T^r_{AV}$, and service rate $\alpha$. However, when generating the exogenous variable $P^r_{AV}$ via PSO, random variations are introduced to these values. Directly replacing the non-linear binary logit model with Eq. (42) may lead to infeasibility because of the integer nature of travel time $T^r_{AV}$, and the bounds on the pricing variables $p^0, p, p^c$ and service rate variable $\alpha$.

To maintain feasibility within the model, we relax Eq. (42) by replacing the equality condition with an inequality, resulting in the relaxed constraints as shown in Constraints (43). This relaxation enables us to determine the minimum prices $p^0, p, p^c$ across all groups that satisfy these constraints.

$$V^r_{AV} \geq \ln P^r_{AV} - \ln(1 - P^r_{AV}) + V^r_B, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{43}$$

Additionally, once $P^r_{AV}$ is predetermined as parameters before optimisation, Constraint (9) becomes linear, allowing demand variable $D^r_{AV}$ to be directly determined using Eq. (4). The mathematical formulations for **M2** under the three pricing strategies combine the following equations:

*Pricing strategy 1: base fare + distance-based fare.* **[M2-1]** Objective function (35), subject to Constraints (2)–(5), (8)–(28), (30), (33), (34), (38)–(41), (43), and the non-negativity constraints (A.52), (A.53), (A.57)–(A.60), (A.63)–(A.78) in Appendix A.

*Pricing strategy 2: distance-based fare.* **[M2-2]** Objective function (35), subject to Constraints (2)–(4), (6), (8)–(28), (30), (33), (34), (38)–(41), (43), and the non-negativity constraints (A.53), (A.57)–(A.60), (A.63)–(A.78) in Appendix A.

*Pricing strategy 3: income class-based fare.* **[M2-3]** Objective function (35), subject to Constraints (2)–(4), (7), (8)–(28), (30), (33), (34), (38)–(41), (43), and the non-negativity constraints (A.54), (A.57)–(A.60), (A.63)–(A.78) in Appendix A.

After solving the model, we obtain optimal values for the price, service rate, and travel time variables. These values are used to update the probability of using SAVs in each group, through Eq. (1). These updated probabilities then serve as new inputs for M2. The iterative process continues until one of the following stopping criteria is satisfied: (1) the difference between the input and output probabilities becomes sufficiently small, as illustrated in Fig. 2; or (2) the maximum number of iterations is reached.

#### 4.2.2. Outer loop: particle swarm optimisation (PSO)

Particle Swarm Optimisation (PSO) is a bio-inspired metaheuristic in which a population of particles explores the search space using two reference points: each particle's personal best position (*pbest*) and the best position found by the swarm (*gbest*). This social-cognitive mechanism guides particles toward promising regions of the search space. However, such a mechanism does not imply guaranteed convergence under all parameter settings. To improve search robustness and reduce the likelihood of premature convergence, we adopt a hybrid PSO-Differential Evolution (DE) scheme, in which DE intensifies the search within a confined neighbourhood of the current *gbest* while helping preserve diversity. Previous studies have shown that PSO-DE hybridisation can improve empirical solution quality and convergence speed (Gad, 2022).

**Step 1: Initialisation**

*Initialise the particle swarm.* Each particle represents a candidate solution, and its position is a probability vector whose entries are $P_{AV}^{r}$, the probability that group $r \in \mathcal{R}^c,\ c \in C$, uses the SAV service. We first generate a candidate set of particles with size equal to twice the target swarm size. After evaluating their fitness, we retain the top-performing half using elitist selection to form the initial swarm. This procedure provides a stronger starting point for the optimisation while preserving diversity among particles.

Instead of random initialisation, each particle's probability vector $\{P_{AV}^{1}, P_{AV}^{2}, \ldots, P_{AV}^{\sum_{c\in C}|\mathcal{R}^c|}\}$ is constructed using a grid-based scheme. For each group $r \in \mathcal{R}^c,\ c \in C$, the probability $P_{AV}^{r}$ has group-specific lower and upper bounds. The lower bound is 0. The upper bound is calculated under the most favourable operating conditions, namely the minimum price, the shortest travel time, and a no-rejection policy ($\alpha = 1$), which together yield the highest travel utility and therefore the highest probability of using the SAV service.

Starting from the lower bounds, we increase each probability component $P_{AV}^{r}$ toward its corresponding upper bound using a uniform step size. For each group $r$, the step size is defined as the allowable range of that probability, i.e., $\frac{\overline{P}_{AV}^{r} - \underline{P}_{AV}^{r}}{2N}$, where $N$ is the target swarm size, $\underline{P}_{AV}^{r}$ is the lower bound, and $\overline{P}_{AV}^{r}$ is the upper bound. This process continues until all components reach their respective upper bounds.

This grid-based initialisation is preferable to random sampling because of Constraint (43). If even one probability $P_{AV}^{r}$ is randomly initialised at an excessively high value, the pricing variables may be forced to remain low, which restricts the feasible region explored by the algorithm. As a result, many randomly generated particles become similar and exhibit poor performance. The grid-based scheme instead produces a diverse and well-distributed set of initial particles that better respects the structure of the feasible region.

*Initialise particle velocities.* Each particle is associated with a velocity vector, which determines how its probability vector is adjusted when the position is updated. In this problem, each velocity component specifies the change applied to the corresponding probability $P_{AV}^{r}$ in the next iteration.

Rather than generating initial velocities randomly, we use relative position information from the initial swarm. This idea is inspired by differential evolution and provides more informed initial search directions. For each particle whose fitness is not the current global best (*gbest*), the initial velocity is defined using the displacement from its current position to the position of *gbest*. For the particle corresponding to *gbest*, the initial velocity is defined using the displacement between its current position and that of a randomly selected *pbest*. These vectors determine the initial search directions; the actual movement magnitudes are governed later by the velocity update rule and the clamping operator. Compared with purely random initialisation, this strategy provides a more meaningful starting search pattern and guides the swarm toward promising regions of the search space.

*Initialise the velocity-clamping operator.* Velocity clamping is used to prevent particles from moving either too quickly or too slowly across the search space. In this study, excessively large velocities may cause large jumps in the probability vector, potentially skipping high-quality solutions or pushing probability values outside their feasible bounds. Conversely, excessively small velocities may lead to negligible changes in the probability values and, hence, little variation in total demand. The clamping operator keeps velocity updates within a reasonable range so that the swarm can explore effectively while maintaining feasible and meaningful solutions.

For each group $r \in \mathcal{R}^c,\ c \in C$, we define velocity-clamping parameters $v_{\max}^{r}$ and $v_{\min}^{r}$. The maximum velocity $v_{\max}^{r}$ is set to 20% of the range between the upper and lower probability bounds of that group. This ensures that every particle's velocity remains within the interval$[-v_{\max}^{r}, v_{\max}^{r}]$. The minimum velocity $v_{\min}^{r}$ is defined as $1/n^{r}$, so that the resulting update corresponds to at least one additional trip choosing SAV or one fewer trip choosing SAV for group $r$. However, this minimum threshold does not need to be enforced for every group in every update. Instead, it is activated only when a position update would otherwise produce no change in aggregate demand across all groups. In that case, a subset of groups is selected randomly, and their velocity components are adjusted to either $v_{\min}^{r}$ or $-v_{\min}^{r}$. This ensures that each particle update induces a meaningful change in total demand.

**Step 2: Main loop (repeated until the stopping criterion is met)**

*Step 2.1 Update velocities and positions.* We employ two search strategies to update particle velocities and positions. The first is a dynamic PSO strategy that combines inertia, cognitive learning, and social learning with stochastic weights. The second is a differential-evolution-inspired strategy designed to strengthen local exploitation around the best solution found so far.

**Strategy 1: Dynamic PSO search strategy**

For each particle, the updated velocity vector $v_{\text{new}}$ is determined by its current velocity vector $V$, the displacement from its current position vector $X$ to its personal-best position vector $pbest$, and the displacement from $X$ to the global-best position vector $gbest$:

$$v_{\text{new}} = w \cdot v + c_1(t) \cdot r_1 \cdot (pbest - X) + c_2(t) \cdot r_2 \cdot (gbest - X), \tag{44}$$

where

$$c_1(t) = \left(c_{1,\min} - c_{1,\max}\right)\frac{t}{t_{\max}} + c_{1,\max}, \tag{45}$$

$$c_2(t) = \left(c_{2,\max} - c_{2,\min}\right)\frac{t}{t_{\max}} + c_{2,\min}. \tag{46}$$

In Eq. (44), $w$ denotes the inertia weight, which controls how much of the previous velocity is retained in the new update. The coefficients $c_1(t)$ and $c_2(t)$ are scalar cognitive and social learning factors at iteration $t$, respectively. The cognitive term $c_1(t)$ encourages each particle to move toward its own best historical position, whereas the social term $c_2(t)$ attracts it toward the best solution found by the swarm. In our setting, these terms determine how strongly the probability vector of each particle is adjusted toward promising regions of the search space. The two learning coefficients vary dynamically over the course of the algorithm: $c_1(t)$ decreases gradually, while $c_2(t)$ increases gradually. This design shifts the search emphasis from early-stage exploration based on individual experience to later-stage convergence toward the global-best solution. Their upper and lower bounds are denoted by $c_{1,\max}$, $c_{1,\min}$, $c_{2,\max}$, and $c_{2,\min}$, respectively. A sensitivity analysis of these parameters is provided in Appendix B. The random variables $r_1$ and $r_2$, each independently drawn from a uniform distribution on $[0, 1]$, introduce stochasticity into the velocity update and help maintain a balance between exploration and exploitation.

After computing the updated velocity vector $v_{\text{new}}$, we apply the velocity-clamping operator to ensure that each velocity component remains within its predefined range and that the resulting update induces variation in total demand when necessary. The particle position vector is then updated as

$$X_{\text{new}} = X + v_{\text{new}}. \tag{47}$$

If any component of $X_{\text{new}}$ falls outside its feasible probability range, it is projected onto the nearest feasible bound.

**Strategy 2: Differential-evolution-inspired search strategy**

This strategy enhances local exploitation around the current global-best solution $gbest$. Specifically, each particle position vector $X$ is updated using the current best solution $X_{gbest}$ and two randomly selected particle position vectors $X_1$, and $X_2$:

$$X_{\text{new}} = X_{gbest} + r_1(X_{gbest} - X) + r_2(X_1 - X_2). \tag{48}$$

Here, $X_{\text{new}}$ is the updated vector in the search space, and $r_1$ and $r_2$ are scalar random variables. Thus, both $X_{gbest} - X$ and $X_1 - X_2$ are displacement vectors. The first term encourages movement toward the best solution found so far, while the second introduces search diversity based on the relative positions of other particles. After the update, each component of $X_{\text{new}}$ is checked against its probability bounds and projected to the nearest feasible boundary if required.

*Step 2.2 Evaluate fitness.* The updated particle positions are then used as inputs to model M2. The resulting objective value of model M2 is taken as the fitness of each particle.

*Step 2.3 Update $pbest$ and $gbest$.* For each particle, we compare the fitness at its new position with the fitness at its current $pbest$. If the new position yields a better fitness value, $pbest$ is updated accordingly. We then compare all updated $pbest$ values and update $gbest$ if any particle achieves a fitness value better than the current global best.

**Step 3: Termination**

The main loop in Step 2 is repeated either for a predetermined number of iterations or until a convergence condition is satisfied, for example when the improvement in $gbest$ remains negligible over several consecutive iterations. The final solution is given by the position of $gbest$.

### *4.3. Parallel Bayesian optimisation with a reformulated MILP model (M3)*

Bayesian Optimisation (BO) has proven to be a powerful tool for exploring the parameter space efficiently (Bergstra et al., 2011; Swersky et al., 2013; Liu et al., 2019). It employs a surrogate model to guide the selection of the next sampling points, with the goal of identifying optimal parameters with minimal evaluations. Instead of resorting to binary variables and the big-M method to reformulate the products of continuous and integer variables in the objective function (29), BO provides a sequential search strategy that allows us to strategically target the best pricing and service rate parameters to maximise the objective function value.

The framework of the parallel BO with the reformulated MILP model (M3) is illustrated in Fig. 3. A more detailed description of the reformulated model M3, the surrogate model, and the acquisition function is provided in the following sections.

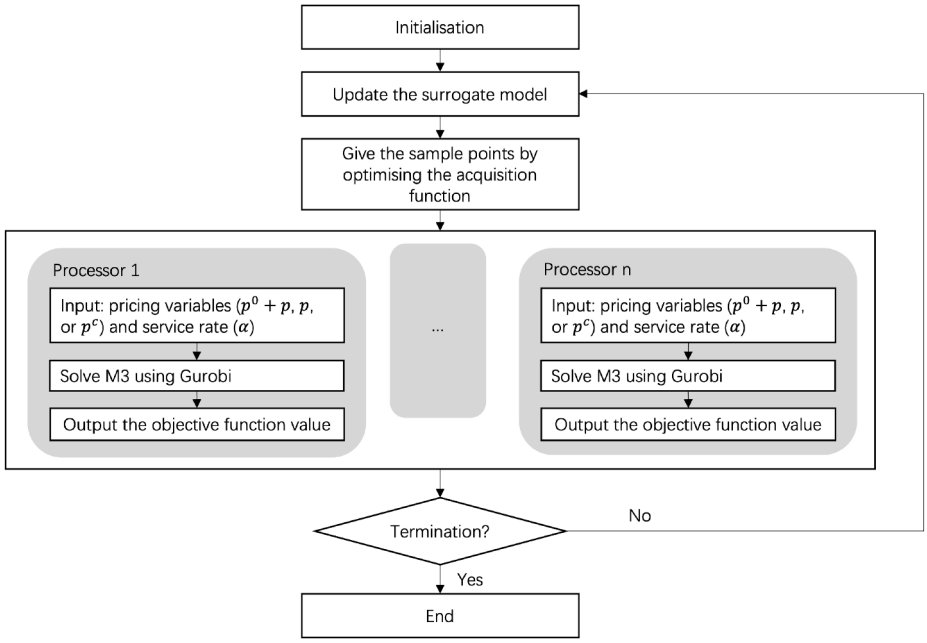



**Fig. 3.** Framework of the Bayesian optimisation with reformulated MILP model (M3).

#### 4.3.1. *Reformulated MILP model (M3)*

By treating the service rate and pricing variables as parameters in Constraint (9) and objective function (29), we reduce the number of binary variables and constraints in Model (M1). Note that we still need to linearise the binary logit model using the outer-inner approximation method. The formulation of the reformulated MILP, referred to as M3, is detailed under the three pricing strategies as follows:

**Pricing strategy 1: base fare + distance-based fare**

**[M3-1]** Objective function (29) with Constraints (2)–(5), (8)–(28), the linearisation-related Constraints (A.29)–(A.45), and the non-negativity constraints (A.59)–(A.78) in Appendix A.

**Pricing strategy 2: distance-based fare**

**[M3-2]** Objective function (29) with Constraints (2)–(4), (6), (8)–(28), the linearisation-related Constraints (A.29)–(A.45), and the non-negativity constraints (A.59)–(A.78) in Appendix A.

**Pricing strategy 3: income class-based fare**

**[M3-3]** Objective function (29) with Constraints (2)–(4), (7), (8)–(28), the linearisation-related Constraints (A.29)–(A.45), and the non-negativity constraints (A.59)–(A.78) in Appendix A.

The parameters that the BO targets under the three pricing strategies are $[p^0, p, \alpha]$, $[p, \alpha]$, and $[p^c, \alpha]$ with $c \in C$, respectively. The bounds of $p^0$, $p$, $p^c$, and $\alpha$ are determined by combining Constraints (38) with Eqs. (5), (6), and (7). BO can only target the values within these ranges.

$$0 \le p^0 \le \max_{c\in C, r\in R^c} \left( \frac{-V_B^r + \ln(2n^r - 1)}{\beta_0^c} - VOT_{AV}^c t_{\text{short}}^r \right) \tag{49}$$

$$0 \le p \le \max_{c\in C, r\in R^c} \left( \frac{-V_B^r + \ln(2n^r - 1)}{\beta_0^c l_{\text{short}}^r} - \frac{VOT_{AV}^c t_{\text{short}}^r}{l_{\text{short}}^r} \right) \tag{50}$$

$$0 \le p^c \le \max_{c\in C, r\in R^c} \left( \frac{-V_B^r + \ln(2n^r - 1)}{\beta_0^c l_{\text{short}}^r} - \frac{VOT_{AV}^c t_{\text{short}}^r}{l_{\text{short}}^r} \right), \qquad c \in C \tag{51}$$

$$0 \le \alpha \le 1 \tag{52}$$

#### 4.3.2. *Surrogate model and acquisition function*

There are two main components in BO: a surrogate model and an acquisition function. The surrogate model is a probabilistic model that serves as a surrogate for the true objective function. The surrogate model is built using historical observations of the input and the objective function values in all the iterations. In this study, we use a Gaussian Process (GP) as the surrogate model, as it is

widely adopted in BO because of its high data efficiency and its ability to quantify predictive uncertainty (Brochu et al., 2010; Liu et al., 2019).

At iteration $t$, the available dataset is

$$\mathcal{D}_t = \{(\mathbf{x}_i, y_i)\}_{i=1}^{t}, \tag{53}$$

where $\mathbf{x}_i$ denotes the $i$-th evaluated input vector and $y_i$ is the corresponding observed objective value obtained by solving the optimisation model at $\mathbf{x}_i$. These observed values are treated as evaluations of an unknown latent objective function $f(\cdot)$, possibly subject to Gaussian observation noise.

A GP prior is placed on the latent function $f(\cdot)$:

$$f(\mathbf{x}) \sim \mathcal{GP}\big(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\big), \tag{54}$$

where $m(\mathbf{x})$ is the mean function and $k(\mathbf{x}, \mathbf{x}')$ is the covariance kernel function. Here, $\mathbf{x}$ and $\mathbf{x}'$ denote any two input points in the search space. The kernel function specifies the correlation between the latent function values at these two input points, namely $f(\mathbf{x})$ and $f(\mathbf{x}')$. In other words, if $\mathbf{x}$ and $\mathbf{x}'$ are close or similar, then their corresponding function values are expected to be similar as well.

Given the observed data $\mathcal{D}_t$, the GP defines a posterior predictive distribution for the latent objective value at a new input point $\mathbf{x}$. This posterior distribution is characterised by a mean function $\mu_t(\mathbf{x})$ and a variance function $\sigma_t^2(\mathbf{x})$, both obtained by conditioning on the previously observed input–output pairs in $\mathcal{D}_t$. In particular, $\mu_t(\mathbf{x})$ is not a simple average over $\mathbf{x}$; rather, it is the GP estimate of the objective value at $\mathbf{x}$ informed by the previously observed objective values $y_1, \dots, y_t$. Accordingly,

$$f(\mathbf{x}) \mid \mathcal{D}_t \sim \mathcal{N}\big(\mu_t(\mathbf{x}), \sigma_t^2(\mathbf{x})\big), \tag{55}$$

where the posterior mean and variance are given by

$$\mu_t(\mathbf{x}) = \mathbf{k}_t(\mathbf{x})^\top [\mathbf{K}_t + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{y}_t, \tag{56}$$

$$\sigma_t^2(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}_t(\mathbf{x})^\top [\mathbf{K}_t + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{k}_t(\mathbf{x}). \tag{57}$$

Here, $\mathbf{y}_t = [y_1, \dots, y_t]^\top$ is the vector of observed objective values, $\mathbf{k}_t(\mathbf{x}) = [k(\mathbf{x}_1, \mathbf{x}), \dots, k(\mathbf{x}_t, \mathbf{x})]^\top$ is the vector of covariances between the new input point $\mathbf{x}$ and the previously evaluated points, and $\mathbf{K}_t = [k(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1}^{t}$ is the covariance matrix of all previously evaluated inputs. As additional observations are collected, the posterior distribution is updated iteratively, typically improving predictive accuracy and reducing uncertainty.

For the kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$, we adopt the Matérn kernel (Frazier, 2018; Liu et al., 2019), which provides a flexible covariance structure and controls the smoothness of the latent function through the parameter $\nu$:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \Big(2\sqrt{\nu}\, \|\mathbf{x}_i - \mathbf{x}_j\|\Big)^{\nu} H_\nu\Big(2\sqrt{\nu}\, \|\mathbf{x}_i - \mathbf{x}_j\|\Big), \tag{58}$$

where $\Gamma(\cdot)$ and $H_\nu(\cdot)$ denote the Gamma function and the modified Bessel function of order $\nu$, respectively, and $\|\mathbf{x}_i - \mathbf{x}_j\|$ denotes the Euclidean distance between the two input points $\mathbf{x}_i$ and $\mathbf{x}_j$ in the search space. In this study, we set $\nu = \frac{5}{2}$, which yields a sufficiently smooth surrogate model and is commonly used in GP-based optimisation (Williams and Rasmussen, 2006; Liu et al., 2019). Readers may refer to Brochu et al. (2010) and Swersky et al. (2014) for further background on GP-based BO.

To determine the next evaluation point, BO uses an acquisition function $\alpha_t(\mathbf{x})$, which quantifies the utility of sampling at $\mathbf{x}$. In this study, we employ the Expected Improvement (EI) criterion (Brochu et al., 2010; Frazier, 2018), defined as

$$\alpha_t^{\mathrm{EI}}(\mathbf{x}) = \mathbb{E}\big[\max(f(\mathbf{x}) - y^+, 0)\big], \tag{59}$$

where $y^+ = \max_{i=1,\dots,t} y_i$ is the best observed objective value so far. Under the GP posterior, EI has the closed-form expression

$$\alpha_t^{\mathrm{EI}}(\mathbf{x}) = (\mu_t(\mathbf{x}) - y^+ - \xi)\Phi(z) + \sigma_t(\mathbf{x})\phi(z), \tag{60}$$

where

$$z = \frac{\mu_t(\mathbf{x}) - y^+ - \xi}{\sigma_t(\mathbf{x})}, \tag{61}$$

and $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function and probability density function of the standard normal distribution, respectively. The parameter $\xi \geq 0$ controls the exploration–exploitation trade-off.

At each iteration, the next query point is selected by maximising the acquisition function:

$$\mathbf{x}_{t+1} = \arg\max_{\mathbf{x} \in \mathcal{X}} \alpha_t^{\mathrm{EI}}(\mathbf{x}). \tag{62}$$

The optimisation model is then evaluated at $\mathbf{x}_{t+1}$, yielding a new observed objective value $y_{t+1}$. The dataset is updated as

$$\mathcal{D}_{t+1} = \mathcal{D}_t \cup \{(\mathbf{x}_{t+1}, y_{t+1})\}. \tag{63}$$

This process is repeated until the predefined maximum number of iterations is reached. The resulting solution is the best observed over all iterations.

### 4.4. Solution feasibility assessment and correction

This section evaluates the feasibility of solutions produced by the proposed methods relative to the original model and introduces a correction procedure to recover feasible solutions when infeasibilities arise.

#### 4.4.1. Solution feasibility assessment

We examine how solutions from the proposed algorithms relate to the original problem formulation and assess their feasibility. Throughout, **P0** denotes the original mixed-integer non-linear programming model, while **P1**, **P2**, and **P3** denote the linearisation-based model, the hybrid PSO method, and the hybrid BO method, respectively. Since **P1**, **P2**, and **P3** each modify **P0** in different ways, it is not immediately clear whether their solutions satisfy the constraints of the original model. The following three propositions clarify this relationship.

**Proposition 1.** *The optimal objective value obtained by method **P1** is an upper bound on the optimal objective function value of model **P0**. Furthermore, **P0** and **P1** attain the same optimal objective value if an optimal solution of **P1** is feasible for **P0**. However, a solution obtained from **P1** is not necessarily feasible for **P0**.*

**Proof.** **P1** is a relaxation of **P0**: the outer-inner approximation of the non-linear terms enlarges the feasible region, so the feasible set of **P0** is contained in that of **P1**. For a maximisation problem, this implies that the optimal value of **P1** is an upper bound on that of **P0**. If an optimal solution of **P1** is also feasible for **P0**, it attains the same objective value in both models and is therefore optimal for **P0** as well. Since the relaxation strictly enlarges the feasible region, solutions of **P1** are not guaranteed to lie in the feasible set of **P0**. □

**Proposition 2.** *Solutions obtained by **P2** are feasible for **P0** if, in the inner loop of **M2**, the updated probabilities of using SAVs equal the input probabilities. However, global optimality cannot be guaranteed.*

**Proof.** In **P2**, the binary logit model is handled outside the optimisation model through an iterative algorithm. If the updated probabilities match the corresponding input probabilities, then the binary logit relation is satisfied. The remaining nonlinear terms in the objective function, arising from products of continuous and integer variables, are treated through exact reformulation. Hence, under this condition, a solution obtained by **P2** is feasible for **P0**. However, because **P2** is a meta-heuristic method, global optimality cannot be guaranteed. □

**Proposition 3.** *Models **P1** and **P3** share the same feasible region and are therefore the same in optimal objective value. However, a solution obtained from **P3** is not necessarily feasible for **P0**.*

**Proof.** Model **P3** explores the same feasible region as **P1**, while differing only in the solution approach. It follows that **P1** and **P3** have the same feasible region and, therefore, the same optimal objective value. Because **P1** is a relaxation of **P0**, this feasible region is not necessarily feasible for **P0**. Hence, a solution obtained from **P3** is not guaranteed to be feasible for **P0**. □

#### 4.4.2. Correction of infeasible solutions

If a solution obtained from **P1**, **P2**, or **P3** is infeasible for the original model, we apply a feasibility recovery procedure to construct a feasible solution while preserving the main strategic decisions of the obtained infeasible solution. The infeasibility mainly arises from the approximation of the non-linear binary logit constraints. Therefore, instead of completely discarding the obtained infeasible solution, we fix some variables obtained from it in the original model, namely, the global service rate $\alpha$ and for all groups of trips $c \in C, r \in \mathcal{R}^c$, the monetary cost $c^r_{AV}$ and the longest travel time $T^r_{AV}$. Once these variables are fixed, for all groups of trips $c \in C, r \in \mathcal{R}^c$, the utility $V^r$, choice probability $P^r_{AV}$, and corresponding SAV demand $D^r_{AV}$ follow from Constraints (2), (1), and (4), respectively. Thus, the nonlinearities arising from the binary logit model in Constraints (1), from Constraints (9) and objective function (29) are also eliminated. The remaining variables, including the number of served requests $S^r$ for all groups of trips $c \in C, r \in \mathcal{R}^c$, fleet size $O$, and the passenger and vehicle flow-related variables such as $F_{i_{t1}j_{t2}}$, and $\hat{F}_{i_{t1}j_{t2}}$ for all $\left(i_{t_1}, j_{t_2}\right) \in \mathcal{G}$, are then re-optimised using the remaining linear part of the original formulation. In this way, the recovered solution satisfies the original constraints while retaining some of the essential strategic choices of the obtained solution.

Fig. 4 illustrates the dependency structure among the main variables in the original model and clarifies why the recovery procedure is valid. The main source of potential infeasibility is the mismatch in the demand calculation: when the approximated demand used in **P1**, **P2** and **P3** deviates substantially from the demand implied by the original binary logit model, the fixed longest travel time $T^r_{AV}$ may become inconsistent with the recomputed demand $D^r_{AV}$, leaving no feasible flow distribution in the remaining linear part of **P0**. The main steps of the recovery procedure are summarised in Algorithm 1.

## 5. Case study of the city of Delft, in the Netherlands

We apply the proposed solution methods to a case study of Delft, in the Netherlands. This section first introduces the application setting in Section 5.1. We then test the methods on a small problem involving only 9 groups of trips to compare the performance of the three proposed methods in Section 5.2. Following this small case study, we extend the application to the entire city of Delft. The optimisation results are presented in Section 5.3.

### 5.1. Application setting

We introduce the application setting from the following perspectives: network description, mobility data, optimisation setting, parameter setting, and computational setting.

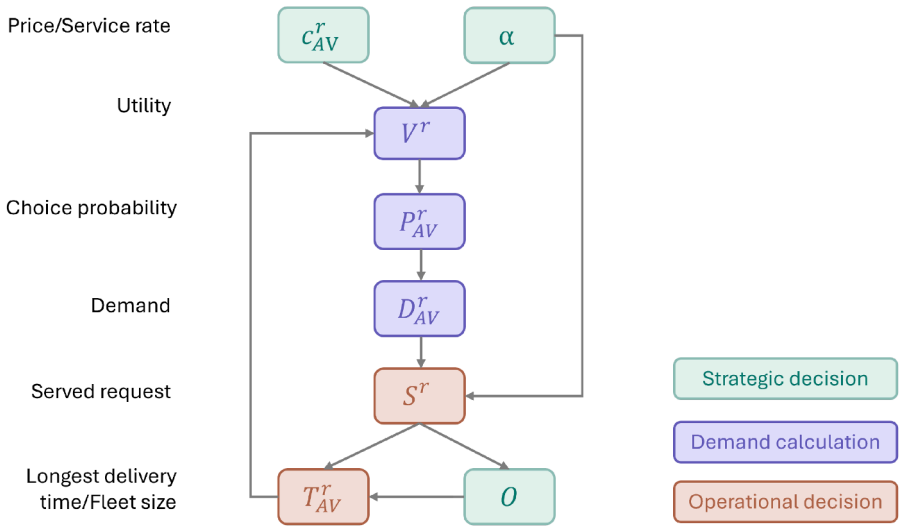



**Fig. 4.** Dependencies among the main variables in the original model.

**Algorithm 1** Repair procedure for infeasible solutions.

**Input:** Solution obtained from **P1**, **P2**, or **P3**
**if** the obtained solution is feasible for the original model **then**
    Return the original solution
**else**
    Fix $c_{AV}^r$, $T_{AV}^r$, and $\alpha$ at the values obtained from **P1**, **P2**, or **P3**
    Compute first $V_{AV}^r$ and then $P_{AV}^r$ using the binary logit model
    Determine $D_{AV}^r$ accordingly
    Re-optimise the remaining decision variables in the original model **P0**, such as $O$, $S^r$, $F_{i_{t1}j_{t2}}$, and $\hat{F}_{i_{t1}j_{t2}}$
    **if** the resulting MILP is feasible **then**
        Return the repaired feasible solution
    **else**
        Report that no feasible solution exists under the fixed values of $c_{AV}^r$, $T_{AV}^r$, and $\alpha$
    **end if**
**end if**

*Network description.* We apply the proposed model and solution methods to a quasi-real case study of Delft, a city in the South-Holland province in the Netherlands, using the simplified road network detailed by Fan et al. (2023). The network consists of 35 nodes and 104 directed links that allow two-way traffic, as depicted in Fig. 5. SAVs can navigate the entire network, but only seven nodes (3, 10, 11, 15, 19, 22, and 27) are designated as free parking depots, marked in red on the map. The road links are designed with capacities of 1600 vehicles for single lanes and 3200 for double lanes per hour, with speed limits of 50 km/h and 70 km/h, respectively. A minimum speed of 5 km/h is enforced on all roads to ensure that each link has a reasonable upper bound on travel time. This assumption is both realistic and essential for constructing the time-space network under congestion, as it prevents excessively long travel times that would otherwise result in a large set of time-space links.

*Mobility data.* We utilise the Dutch mobility dataset (MON 2007/2008) (Correia and Van Arem, 2016; Liang et al., 2020; Fan et al., 2023) to extract detailed travel demand for bicycles, cars, and taxis between 7 am and 10 am. To represent these data effectively for the morning peak hour, we evenly distribute these trips over an hour, creating a total of 2933 trips. These are further aggregated into 45 groups based on similarity in trip characteristics.

*Optimisation setting.* The optimisation considers a one-hour morning peak, divided into 24 time steps of 2.5 minutes each. To accommodate the trips from the parking depots, an additional 5 time steps are included before the peak, resulting in a total of 29 time steps for the complete optimisation period.

*Parameter setting.* The parameters related to the network configuration, demand characteristics, and SAV operational dynamics in this case study are outlined here. The maximum travel time $t_{ij}^{\max}$ for SAVs to travel through link $(i,j) \in \mathcal{L}$, is computed by dividing the length of each road link by the minimum travel speed of 5 km/h, ensuring that the vehicles adhere to speed limits. Conversely, the minimum travel time $t_{ij}^{\min}$ for SAVs to travel through link $(i,j) \in \mathcal{L}$, is determined by dividing the road link length by the respective

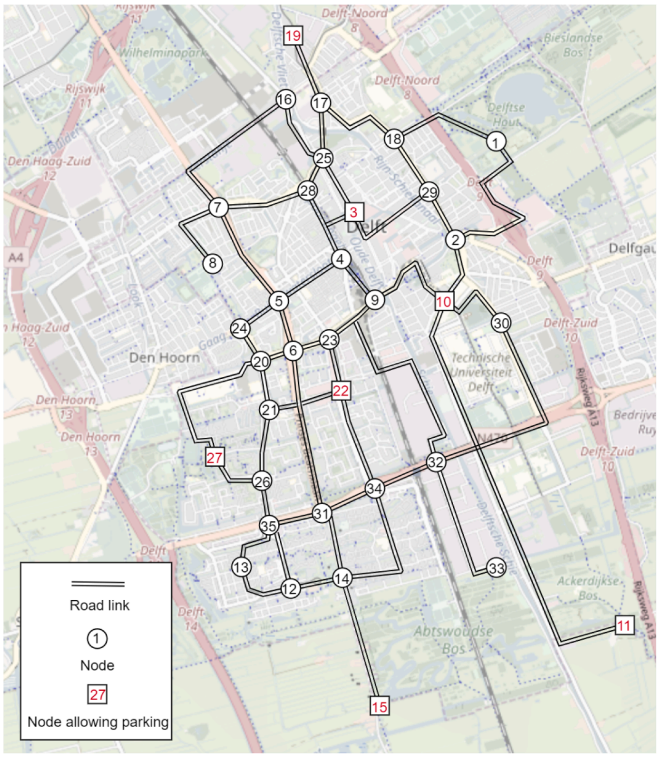



**Fig. 5.** Simplified road network of Delft used in the case study.

maximum speeds of 50 km/h or 70 km/h. Due to the time-space network model used in this study, the minimum travel time on each road link cannot fall below one time step. This ensures realism in the network model, as it prevents scenarios where a vehicle could traverse a link in zero time.

To obtain the shortest travel distances and times, $l^r_{\text{short}}$ and $t^r_{\text{short}}$ for $c \in C, r \in \mathcal{R}^c$, we use the shortest path algorithm assuming vehicles can travel at free-flow speeds. The travel time for bicycles $T^r_B$ with $c \in C, r \in \mathcal{R}^c$ is determined by dividing the shortest path length by the average cycling speed of 12.4 km/h, reflective of typical Dutch cycling habits as noted by BicycleDutch (2018).

We set the logit model base parameter $\beta_0$ at 0.1 (Fan et al., 2023). The VOTT for using an SAV, $VOT^c_{AV}$ with $c \in C$, is set at 6.6, 4.6, and 3.8 euros per hour for high, middle, and low-income travellers, respectively, as estimated by Kolarova et al. (2019). Similarly, the VOTT for using a bicycle, $VOT^c_B$ with $c \in C$, is set at 24.9, 17.3, and 14.1 euros per hour for the respective income groups, following the same source.

The operational cost of SAVs, $c_o$, is determined to be 0.32 euros per km, based on the methodology from Bösch et al. (2018), while the depreciation cost $c_f$ is set at 1.2 euros per vehicle per hour, as identified in Fan et al. (2022). Additionally, the delay penalty $c_d$ is fixed at 0.2 euros per minute, following Liang et al. (2020). Finally, the parameters $a$ and $b$ in the BPR function are set to 2 and 4, respectively, aligning with the values suggested by Van Essen and Correia (2019).

*Algorithm parameter setting.* We first present the parameter settings for **P2**, introduced in Section 4.2. The PSO population size $N$ is set to 12. During the first 20 iterations, the dynamic PSO search strategy is applied; thereafter, the algorithm switches to the DE search strategy and continues until the 30-minute time limit is reached. The inertia weight $w$ is set to 0.2. In addition, the cognitive and social coefficients are specified as $c_{1,\max} = 0.8$, $c_{2,\max} = 0.8$, $c_{1,\min} = 0.1$, and $c_{2,\min} = 0.1$. The inner loop terminates when either the number of iterations reaches 5 or the relative change in each element of the probability vector between two consecutive iterations is no more than 5%. A sensitivity analysis of these parameters is provided in Appendix B.

For **P3**, introduced in Section 4.3, the number of evaluated individuals in each iteration is set to 12. Since **P2** and **P3** are both heuristic methods, they are run under the same 30-minute time limit to enable performance comparison under an equal computational budget. In contrast, **P1** is an exact method and, in the small case study, is solved until solver termination without imposing the same time limit. For the Delft case study, the computational time limit is set to 48 hours for all three methods.

*Computational setting.* The proposed algorithms were implemented in Python 3.7 and the proposed MILP models were solved using Gurobi 9.5.2 on DelftBlue Supercomputer (DHPC , 2024).

**Table 2**
Comparison of optimisation results in the small case study using the three proposed methods, together with corrected solutions where needed.

| | Pricing strategy 1 | | | Pricing strategy 2 | | | Pricing strategy 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | P1 | P2 | P3 | P1 | P2 | P3 | P1 | P2 | P3 |
| Obj. value | 313.27 | 310.09 | 309.18 | 291.43 | 290.48 | 286.29 | 319.69 | 315.33 | 306.55 |
| Corrected obj. value | 303.97 | N/A | N/A | 288.34 | N/A | N/A | 311.51 | N/A | 304.10 |
| Computational time (s) | 1040 | 1800 | 1800 | 1537 | 1800 | 1800 | 4340 | 1800 | 1800 |
| MILP gap (%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Price (€) | $p^0 = 4.93$<br>$p = .64$ | $p^0 = 4.56$<br>$p = .72$ | $p^0 = 4.66$<br>$p = .69$ | $p = 1.59$ | $p = 1.59$ | $p = 1.56$ | $p^{low} = 1.35$<br>$p^{mid} = 1.43$<br>$p^{high} = 1.76$ | $p^{low} = 1.34$<br>$p^{mid} = 1.42$<br>$p^{high} = 1.79$ | $p^{low} = 1.22$<br>$p^{mid} = 1.42$<br>$p^{high} = 1.85$ |
| Service rate (%) | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 99.63 |
| Fleet size (vehicles) | 93 | 92 | 92 | 114 | 114 | 118 | 127 | 124 | 125 |
| Corrected fleet size (vehicles) | 90 | N/A | N/A | 113 | N/A | N/A | 124 | N/A | 128 |

### 5.2. Small case study

Before applying the three proposed solution methods to the real-life, city-sized case study of Delft, we first compare their performance on a smaller case study that includes only 9 groups of trips. These groups are randomly selected from the Dutch mobility dataset (MON 2007/2008). By testing the three proposed solution algorithms on this smaller case study, we can ensure that all the involved MILP models can be solved to optimality (resulting in a MILP gap of 0 for all methods in Table 2). This approach facilitates a more effective comparison of the performance of the three proposed solution algorithms.

Table 2 reports the optimisation results of **P1**, **P2**, and **P3** under three pricing strategies. The objective value is used as the main indicator of solution quality. For **P2** and **P3**, multiple MILP models are solved during the search procedure; thus, the table reports the MILP gap and decision variable values associated with the best solution found. To ensure a fair comparison with respect to the original problem **P0**, corrected objective values are computed using Algorithm 1 whenever the obtained solution is infeasible for **P0**.

Overall, all three methods are able to identify good-quality solutions across all pricing strategies within a short computational time. Among them, **P2** achieves the best objective function values in this small case study. Besides, the corresponding key decision variables, namely the price, service rate, and (corrected) fleet size, are generally similar across the three methods. These results indicate that all three proposed methods are effective for the problem considered and are capable of producing solutions of comparable quality.

We further present the current best objective function values identified by the three algorithms over computational time in Fig. 6. Specifically, Fig. 6(a)–(c) compare their performance under pricing strategies 1, 2, and 3, respectively. These subfigures show that **P3** reaches its best solution in the shortest time among the three algorithms. At the beginning, it spends some time exploring the feasible region, which results in several solutions with relatively low objective function values. However, it then quickly identifies the most promising region and improves the objective value rapidly. For **P2**, we report only the best solution found in each generation in these figures. Due to the elitism strategy used in the initial population generation, **P2** is able to identify a high-quality solution already in the first generation. Overall, all three proposed methods are able to find good-quality solutions within a short computational time, after which their performance stabilises.

From a theoretical perspective, in **P1**, the approximation accuracy of the logarithmic terms can be improved by increasing the number of breakpoints in the outer–inner approximation, although this comes at the cost of substantially higher computational effort. To examine how a tighter approximation affects the solution quality before and after correction, **P1** is solved with 60 breakpoints on the small instance with 9 groups. Table 3 reports the corresponding results. The computational time increases considerably relative to the standard setting, exceeding 10,000 seconds for all three pricing strategies and reaching approximately 118,000 seconds under pricing strategy 3. This confirms that such a fine-grained approximation is computationally impractical for larger instances.

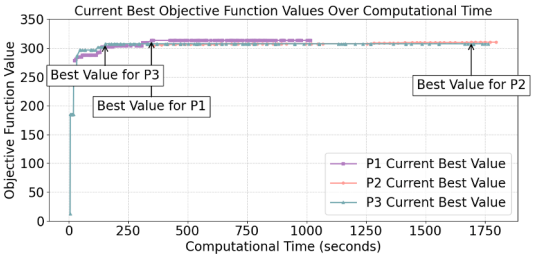
The objective function value worsens as the number of breakpoints increases, compared to P1 in Table 2. This is because additional breakpoints tighten the feasible region, leading to a more constrained problem. Comparing the corrected results in Table 3 with those in Table 2 shows that **P1** with 60 breakpoints still does not outperform **P2**. However, for pricing strategies 2 and 3, it does yield better corrected solutions than **P3**. These results suggest that 60 breakpoints remain insufficient to surpass **P2**, despite the substantially higher computation time they require.

### 5.3. Delft case study

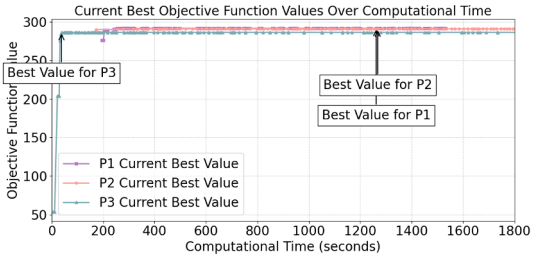
In this section, we present the optimisation results of the Delft case study introduced in Section 5.1 and analyse the interplay among the endogenous decision variables. This analysis includes the impact of the three pricing strategies, income class analysis, congestion effects.

#### 5.3.1. Comparative analysis of algorithm performance

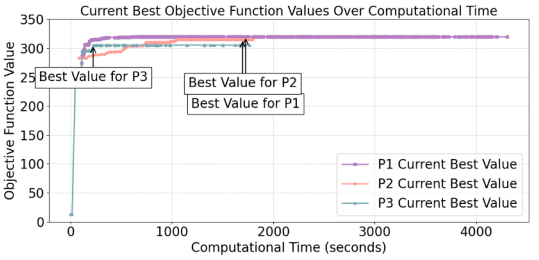
Due to the complexity and scale of the problem, it is challenging to solve all the MILP models to optimality. Therefore, we established specific stopping rules for different MILP models. For **P1**, we set the time limit for solving the MILP model (M1) at 48 hours. For **P2**, when solving the MILP model (M2), we first allow the solver to run for 20 minutes, during which it searches for the




(a) Performance comparison under pricing strategy 1




(b) Performance comparison under pricing strategy 2




(c) Performance comparison under pricing strategy 3

**Fig. 6.** Comparison of algorithm performance under the three pricing strategies in the small case study.

**Table 3**
Optimisation results of **P1** with an increased number of breakpoints in the small case study (9 groups).

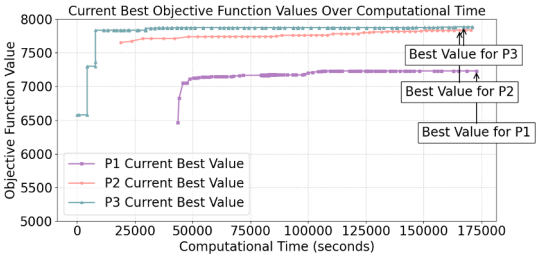
| | Pricing strategy 1 | Pricing strategy 2 | Pricing strategy 3 |
|---|---|---|---|
| Number of breakpoints | 60 | 60 | 60 |
| Obj. value | 311.71 | 290.93 | 318.0 |
| Corrected obj. value | 306.22 | 287.11 | 310.51 |
| Computation time (s) | 10,231 | 10,300 | 118,079 |
| Price (€) | $p^0 = 4.95,\ p = 0.62$ | $p = 1.60$ | $p^{\text{low}} = 1.34,\ p^{\text{mid}} = 1.42,\ p^{\text{high}} = 1.76$ |
| Service rate (%) | 1.0 | 1.0 | 1.0 |
| Fleet size (vehicles) | 94 | 112 | 127.0 |
| Corrected fleet size (vehicles) | 92 | 111 | 123.0 |

**Table 4**
Comparison of optimisation results in the Delft case study using the three proposed methods, together with feasibility assessment and corrected results where applicable.

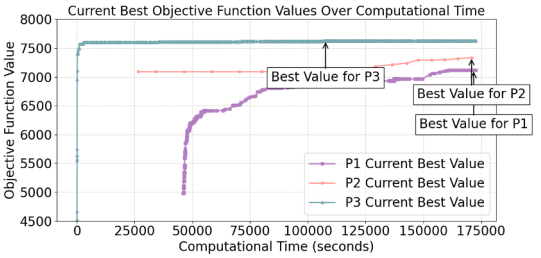
| | Pricing strategy 1 | | | Pricing strategy 2 | | | Pricing strategy 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | P1 | P2 | P3 | P1 | P2 | P3 | P1 | P2 | P3 |
| Obj. value | 7230.91 | 7840.61 | 7887.96 | 7113.40 | 7351.23 | 7628.11 | 9143.37 | 9294.53 | 9352.60 |
| Corrected obj. value | 7173.94 | N/A | 7872.44 | 7257.12 | 7353.07 | 7561.70 | 9096.43 | 9311.34 | 9341.36 |
| MILP gap (%) | 92.8 | 0 | 0.31 | 46.3 | 0.30 | 0.02 | 45.1 | 2.18 | 5.47 |
| Price (€) | $p^0 = 1.16$ $p = 2.54$ | $p^0 = 1.93$ $p = 1.01$ | $p^0 = 1.55$ $p = 1.08$ | $p = 1.09$ | $p = 1.16$ | $p = 1.33$ | $p^{low} = 0.99$ $p^{mid} = 1.17$ $p^{high} = 1.45$ | $p^{low} = 0.96$ $p^{mid} = 1.12$ $p^{high} = 1.52$ | $p^{low} = 0.94$ $p^{mid} = 1.12$ $p^{high} = 1.59$ |
| Service rate (%) | 100 | 98.71 | 100 | 99.78 | 95.28 | 97.86 | 95.72 | 95.48 | 95.47 |
| Fleet size (vehicles) | 596 | 1005 | 966 | 1296 | 1122 | 852 | 1193 | 1249 | 1227 |
| Corrected fleet size (vehicles) | 595 | N/A | 964 | 804 | 1123 | 850 | 1186 | 1251 | 1224 |

best possible solution. After these 20 minutes, we check the current optimality gap. If the gap is larger than 3%, the solver continues. It stops when either the gap reaches 3% or the total runtime reaches 40 minutes. Similarly, for **P3**, we start by letting the solver run for 30 minutes to search for the best possible solution. If, at the end of this period, the optimality gap is larger than 3%, the solver continues. It stops when either the gap reaches 3% or the total runtime reaches 60 minutes. For comparison purposes, both the second and third solution algorithms are run to a 48-hour time limit. We believe that these times can be used in such a strategic context.

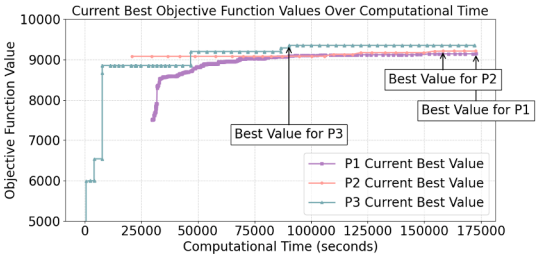
Table 4 reports the optimisation results for the Delft case study under the three pricing strategies. For **P1**, substantial MILP gaps remain when the 48-hour time limit is reached. In particular, under pricing strategy 1, the MILP gap is as high as 92.8%, while under




(a) Performance comparison under pricing strategy 1




(b) Performance comparison under pricing strategy 2




(c) Performance comparison under pricing strategy 3

**Fig. 7.** Comparison of algorithm performance under three pricing strategies in the Delft case study.

**Table 5**
Optimisation results of the Delft case study.

| | Pricing strategy 1 | Pricing strategy 2 | Pricing strategy 3 |
|---|---|---|---|
| Total revenue (€) | 14211.19 | 12919.58 | 17560.34 |
| Total depreciation cost (€) | 1156.80 | 1020.00 | 1468.80 |
| Total operational cost (€) | 4441.45 | 3764.87 | 5645.68 |
| Total delay penalty (€) | 740.50 | 573.00 | 1104.50 |
| Total demand for SAVs (trips) | 1302 | 1263 | 1723 |
| SAV demand share (%) | 44.39 | 43.06 | 58.74 |
| Percentage of satisfied demand (%) | 100.00 | 97.86 | 95.47 |
| Average price per trip (€) | 10.91 | 10.45 | 10.67 |
| Average delay per trip (min) | 1.14 | 0.93 | 1.34 |
| Average delivery time per trip (min) | 6.63 | 5.99 | 6.73 |
| SAVs total travel distance (km) | 13879.53 | 11765.23 | 17642.74 |
| SAVs total delivery distance (km) | 12143.63 | 10227.29 | 15526.54 |
| SAVs total relocate distance (km) | 1735.91 | 1537.94 | 2116.19 |
| Percentage of time parked (%) | 7.31 | 7.08 | 6.53 |

pricing strategies 2 and 3 the gaps are lower, at 46.3% and 45.1%, respectively. This indicates that **P1** becomes computationally challenging for the larger instance.

Comparing the corrected objective values obtained by the three solution methods, **P2** and **P3** outperform **P1** under all three pricing strategies, with **P3** achieving the best overall corrected objective values. Fig. 7 further shows the evolution of the best-found solutions over computational time up to the 48-hour time limit. We can see that both **P2** and **P3** are able to identify high-quality solutions in a much shorter time than **P1**. **P1** takes a long time to find its first feasible solution, and this initial solution is of relatively low quality compared with those found by the other two methods. Although the objective value improves more rapidly afterwards, its overall performance remains weaker. By contrast, **P2** starts from a reasonably good solution, due to the grid search approach and the elitism strategy used in the initial population generation. **P3** exhibits a similar pattern to that observed in the small case study: it first explores the feasible region through several lower-quality trials, and then quickly focuses on the most promising region.

### *5.3.2. Optimisation results*

From the previous section, we observe that **P3** delivers the best performance for the Delft case study. In this section, we analyse the optimisation results obtained using this method. The corresponding results are reported in Table 5.

We first examine the three pricing strategies and their impact on SAV demand and operator performance. The three strategies lead to markedly different revenue and costs. Among them, pricing strategy 3 (income class-based pricing), yields the highest revenue for the SAV operator, as shown in Table 5. It also attracts the highest SAV demand, with 1723 trips, compared to 1302 and 1263 trips

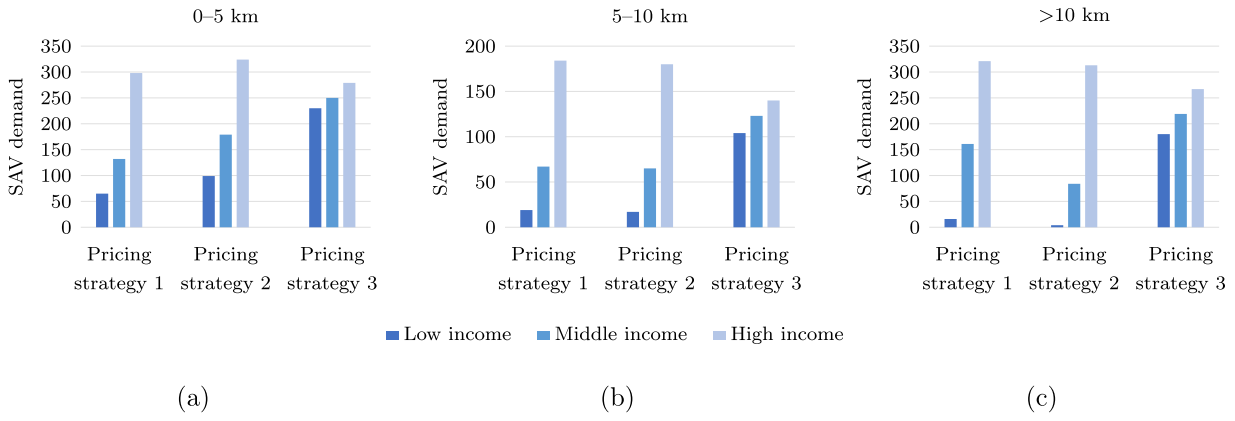



**Fig. 8.** SAV demand for all income classes with trip lengths of (a) 0–5 km, (b) 5–10 km, and (c) >10 km.

under pricing strategies 1 and 2, respectively. This result suggests that differentiated pricing by income class is the most effective strategy for increasing both ridership and operator revenue in the Delft case study.

A more detailed comparison by income group is provided in Fig. 8. Under pricing strategy 3, high-income travellers are charged the highest distance-based fares, followed by middle-income and low-income travellers. Since higher-income travellers are less price-sensitive, the SAV service remains attractive to them even at higher prices. At the same time, pricing strategy 3 increases demand among low- and middle-income travellers compared with pricing strategies 1 and 2. As a result, the differences in SAV demand across income classes become smaller under pricing strategy 3. Interestingly, the demand from high-income travellers decreases slightly under pricing strategy 3 relative to the other two strategies. This can be explained by the higher fares charged to this group and by their higher value of travel time (VOTT), which makes them more sensitive to the increased travel times caused by congestion.

Pricing strategy 1 includes a base fare for every trip, regardless of travel distance. This pricing structure discourages the use of SAVs for short trips. As shown in Fig. 8, for trips shorter than 5 km, pricing strategy 2 generates higher SAV demand than pricing strategy 1 across all income classes. A similar pattern is observed under pricing strategy 3 for low- and middle-income travellers, though not for high-income travellers, who face higher fares under this strategy. In this sense, pricing strategy 1 is more supportive of active transport for short-distance travel. At the same time, Table 5 shows that pricing strategy 1 yields the highest average revenue per trip among the three strategies. Despite attracting a demand level similar to that under pricing strategy 2, it generates higher total revenue for the operator.

Pricing strategy 2 relies only on a distance-based fare. It attracts a demand level close to that of pricing strategy 1, but the operator serves a slightly smaller proportion of demand. As a result, this strategy requires the smallest fleet and leads to the lowest depreciation cost, operational cost, and delay penalty among the three strategies.

Congestion effects are closely related to fleet size, road capacity, and the spatial distribution of demand. Pricing strategy 2 leads to the least congestion, as reflected in the lowest average delay per trip and the smallest total travel distance. By contrast, pricing strategy 3 uses the largest fleet and attracts the most SAV users, which increases both total travel distance and delay. Nevertheless, it still achieves the highest objective value and the highest revenue. In addition, although the fleet size under pricing strategy 3 is the largest among the three strategies, the percentage of time SAVs are parked is the lowest, indicating higher fleet utilisation. Overall, from the operator's perspective, pricing strategy 3 is the most attractive option, as it achieves the best balance between demand attraction and profitability, despite slightly higher congestion-related impacts.

Several more conclusions can be drawn from Fig. 8. Firstly, for short trips of less than 5 km, individuals from low and middle-income classes are particularly sensitive to the pricing strategy, and a fixed base fare is not appealing to them. Secondly, as trip length increases, the demand difference between low, middle, and high-income classes becomes more pronounced. Higher-income individuals are more likely to choose the SAV service due to the higher VOTT of using bicycles. Lastly, the willingness to use the SAV service increases with income, indicating that higher-income individuals are more inclined to opt for SAVs.

### *5.3.3. Income class analysis*

In this study, we explicitly incorporate income-related differences in both cost sensitivity and value of travel time. Since demand is modelled as a function of generalised cost, which combines the monetary price and the value of time, travellers from different income groups are expected to respond differently to pricing and travel time due to variations in their marginal utilities of money and time. Without accounting for these income-based variations, demand modelling would risk oversimplifying the behavioural response to pricing and service attributes. Beyond incorporating income heterogeneity in demand modelling, we also introduce income-based pricing as a potential strategy. From a social perspective, income-based pricing can enhance transport equity by improving accessibility for low-income travellers. From an operational perspective, it offers a mechanism to balance equity and efficiency: fare differentiation moderates excessive demand from high-income users, ensures fair access for lower-income groups, and supports financial sustainability. As policymakers increasingly emphasise equity-oriented mobility design, income-based pricing offers a practical means of embedding fairness considerations into the pricing of SAVs, MaaS platforms, and other on-demand services.

**Table 6**
Comparison of optimisation results: three income classes versus a single income class.

| | 3 classes | | 1 class | |
|---|---|---|---|---|
| | Pricing strategy 1 | Pricing strategy 2 | Pricing strategy 1 | Pricing strategy 2 |
| Price (€) | $p^0 = 1.55$, $p = 1.08$ | $p = 1.33$ | $p^0 = 0.40$, $p = 1.12$ | $p = 1.13$ |
| Fleet size (vehicles) | 964 | 850 | 1098 | 1253 |
| Total revenue (€) | 14211.19 | 12919.58 | 14906.65 | 16349.22 |
| Total depreciation cost (€) | 1156.80 | 1020.00 | 1317.60 | 1503.60 |
| Total operational cost (€) | 4441.45 | 3764.87 | 4967.26 | 5825.63 |
| Total delay penalty (€) | 740.50 | 573.00 | 2104 | 1409 |
| Total demand for SAVs (trips) | 1302 | 1263 | 1552 | 1747 |
| SAV demand share (%) | 44.39 | 43.06 | 52.92 | 59.56 |
| Percentage of satisfied demand (%) | 100.00 | 97.86 | 96.01 | 100 |
| Average price per trip (€) | 10.91 | 10.45 | 10 | 9.36 |
| Average delay per trip (min) | 1.14 | 0.93 | 1.41 | 1.61 |
| Average delivery time per trip (min) | 6.63 | 5.99 | 6.81 | 6.90 |
| SAVs total travel distance (km) | 13879.53 | 11765.23 | 15522.70 | 18205.10 |
| SAVs total delivery distance (km) | 12143.63 | 10227.29 | 13588.47 | 15849.96 |
| SAVs total relocate distance (km) | 1735.91 | 1537.94 | 1934.24 | 2355.15 |
| Percentage of time parked (%) | 7.31 | 7.08 | 7.29 | 6.92 |

In this section, we present several quantitative analyses focusing on two dimensions: (1) the impact of income class segmentation, and (2) the impact of income distribution. These analyses collectively demonstrate the necessity of accounting for income-based heterogeneity and provide insights into the selection of appropriate pricing strategies under different population compositions. However, it is not the objective of this paper to determine optimal income classes to be applied, nor to propose a method to generate such classes.

*Impact of income-class segments.* To quantitatively support our argument, we conducted additional experiments using a version of the model that does not distinguish between income classes. In this simplified setting, heterogeneity arising from income differences was excluded from Eqs. (2) and (3). Instead, all travellers were assumed to share the parameters corresponding to the middle-income group. In addition, income-based pricing was not incorporated in this version of the model. As a result, Pricing Strategy 3 becomes equivalent to Pricing Strategy 2, and only Pricing Strategy 1 (base fare + distance-based fare) and Pricing Strategy 2 (distance-based fare) are effectively distinct in the analysis.

As shown in Table 6, the optimisation results differ substantially when income-based heterogeneity is ignored. In particular, the average revenue per trip is higher in the three-class case than in the single-class case. This is because the heterogeneous specification captures the higher willingness to pay for high-income travellers, allowing the operator to charge higher fares on average. Although the average revenue per trip is higher in the three-class case, the overall SAV demand is lower, which in turn leads to a smaller optimal fleet size, as well as lower operating costs and delay penalties.

By contrast, the single-class specification assumes that all users behave like an average traveller, thereby smoothing out differences in willingness to pay and sensitivity to travel time. This leads to lower average fares per trip, which stimulates higher demand and requires a larger fleet, resulting in higher operating costs. Consequently, the single-class specification tends to overestimate the market uptake of SAV services and may therefore provide a misleading picture for transport planners and service operators. Overall, incorporating income-based heterogeneity produces results that are more behaviourally realistic and operationally informative.

The number of income classes is also an important modelling choice. In this study, we consider three income groups—low, middle, and high income—as this is a common segmentation in income-based transport studies (Kolarova et al., 2019; Dixit and Sivakumar, 2020; Liao et al., 2024). This choice provides a reasonable balance between behavioural realism and computational tractability. A larger number of income classes could capture finer behavioural differences and potentially support more targeted pricing strategies, but it would also increase the computational burden of the optimisation model. More importantly, such an extension would require additional group-specific behavioural parameters, which are not currently available. Without reliable estimates of these parameters, it is difficult to justify a finer segmentation in a meaningful way. Since the main focus of this paper is on optimisation rather than behavioural estimation, we leave the investigation of a more detailed income segmentation to future research.

*Impact of income distribution.* The income distribution, meaning the proportion of people in each income class, can influence the pricing strategy. In the base scenario, we assume that the low-, middle-, and high-income classes each account for an equal share of requests. In this section, we investigate how alternative income distributions affect the resulting pricing strategies.

National-level studies of income class shares in European countries suggest that the middle class typically represents between 51% and 75% of the population in EU member states (Eurofound, 2024), while the low- and high-income groups generally make up smaller shares. To explore alternative income distributions that may occur in specific regions, we test three additional scenarios:

(1) Middle-income dominant, comprising 20% low-, 60% middle-, and 20% high-income individuals;
(2) Low-income dominant, comprising 50% low-, 40% middle-, and 10% high-income individuals;
(3) High-income dominant, comprising 15% low-, 35% middle-, and 50% high-income individuals.

**Table 7**
Comparison of pricing strategies across income-dominant scenarios.

| | Pricing strategy 1 | Pricing strategy 2 | Pricing strategy 3 |
|---|---|---|---|
| **Objective function value** | | | |
| Low-income dominant scenario | 6308.71 | 6085.52 | 6975.02 |
| Middle-income dominant scenario | 7786.80 | 7448.80 | 8371.90 |
| High-income dominant scenario | 9198.63 | 9331.77 | 10778.73 |
| **Price (€)** | | | |
| Low-income dominant scenario | $p^0 = 0.79,\ p = 1.02$ | $p = 1.05$ | $p^{low} = 0.95,\ p^{mid} = 1.13,\ p^{high} = 1.56$ |
| Middle-income dominant scenario | $p^0 = 1.46,\ p = 1.02$ | $p = 1.16$ | $p^{low} = 1.15,\ p^{mid} = 1.13,\ p^{high} = 1.52$ |
| High-income dominant scenario | $p^0 = 3.28,\ p = .95$ | $p = 1.46$ | $p^{low} = 1,\ p^{mid} = 1.21,\ p^{high} = 1.56$ |
| **Fleet size (vehicles)** | | | |
| Low-income dominant scenario | 971 | 1122 | 1172 |
| Middle-income dominant scenario | 1084 | 1140 | 1074 |
| High-income dominant scenario | 1069 | 909 | 1167 |
| **Service rate (%)** | | | |
| Low-income dominant scenario | 98.28% | 95.75% | 100% |
| Middle-income dominant scenario | 100% | 99.74% | 100% |
| High-income dominant scenario | 99.74% | 100% | 100% |

**Table 8**
Optimisation results for the scenario without congestion effects.

| | Pricing strategy 1 | Pricing strategy 2 | Pricing strategy 3 |
|---|---|---|---|
| Price (€) | $p^0 = 1.13,\ p = 0.96$ | $p = 1.08$ | $p^{low} = 0.89,\ p^{mid} = 1.08,\ p^{high} = 1.53$ |
| Fleet size (vehicles) | 1357 | 1385 | 1505 |
| Total revenue (€) | 17245.38 | 17117.39 | 20809.38 |
| Total depreciation cost (€) | 1628.40 | 1662 | 1806 |
| Total operational cost (€) | 5962.55 | 5975.01 | 6606.49 |
| Total delay penalty (€) | 0 | 0 | 0 |
| Total demand for SAVs (trips) | 1826 | 1957 | 2065 |
| SAV demand share (%) | 62.26% | 66.72% | 70.41% |
| Percentage of satisfied demand (%) | 100 | 100 | 100 |
| Average price per trip (€) | 9.45 | 8.78 | 10.12 |
| Average delay per trip (min) | 0 | 0 | 0 |
| Average delivery time per trip (min) | 5.49 | 5.22 | 5.41 |
| SAVs total travel distance (km) | 18632.98 | 18671.91 | 20645.27 |
| SAVs total delivery distance (km) | 16168.52 | 16213.48 | 17950.25 |
| SAVs total relocate distance (km) | 2464.46 | 2458.43 | 2695.02 |
| Percentage of time parked (%) | 7.33 | 7.36 | 7.64 |

These scenarios are intended to be plausible rather than strictly nationally representative. They may reflect suburban contexts or high-income neighbourhoods within urban areas. Table 7 compares the pricing strategies across income-dominant scenarios. The objective function values and key decision variables, including price, fleet size, and service rate, are presented for comparison. As shown in the table, the high-income-dominant scenario consistently achieves the highest objective value under all three pricing strategies. This aligns with the pricing decisions listed below: under Pricing Strategies 1 and 2, the high-income-dominant scenario always sets the highest price for SAV users, reflecting the lower cost sensitivity of high-income travellers. However, under Pricing Strategy 3, the prices for low-, middle-, and high-income users remain relatively stable across scenarios. This suggests that Pricing Strategy 3 performs robustly in zones with different income compositions. In contrast, when a universal price is applied to all income groups, the pricing strategy must be tailored to the actual income distribution of the zone.

### *5.3.4. Congestion analysis*

To illustrate the importance of accounting for congestion effects in pricing and fleet management decisions, we also evaluate scenarios in which congestion is entirely disregarded. In these scenarios, all vehicles travel at free-flow speeds, and link travel times remain fixed regardless of traffic volume. Table 8 presents the optimisation results obtained under this no-congestion assumption.

When congestion effects are removed, all trips are completed at the shortest possible travel time, resulting in no delay penalties in the objective function. However, such consistently fast travel is unlikely to occur in real-world conditions. This idealised setting increases users' willingness to adopt SAV services, as travel times become shorter than in the congestion-aware scenario. As shown in Table 8, the resulting SAV demand is considerably higher than the corresponding results in Table 5, which in turn requires a larger fleet. The optimal prices across all three pricing strategies are also lower in the no-congestion case, which attracts more demand for SAV services. Overall, neglecting congestion leads to unrealistic travel-time assumptions and consequently overestimates both demand and fleet requirements, producing pricing and fleet management decisions that would not be viable in practice.

## 6. Conclusion and future work

This study investigates optimal pricing and fleet-management decisions for an SAV service provider in future urban mobility systems. In the proposed setting, travellers choose between SAVs and bicycles, reflecting urban trends towards car-free environments. Travellers submit SAV requests through a platform, which decides whether to accept or reject them based on operational profitability. To analyse this problem, we formulate a Mixed-Integer Nonlinear Programming (MINLP) model that captures demand-supply interactions, congestion effects, and heterogeneous traveller preferences. Three pricing strategies are considered: a base fare plus a distance-based fare, a distance-based fare only, and an income class-based fare.

To solve the proposed MINLP, we develop three solution methods. The first method reformulates the problem through linearisation and solves it using a commercial solver. The second combines Particle Swarm Optimisation (PSO) with an iterative MILP-based framework. The third integrates Bayesian Optimisation (BO) with a reformulated MILP model to guide the search over the most promising solutions. Since the proposed methods solve approximated or reformulated versions of the original problem, we further assess the feasibility of the obtained solutions with respect to the original MINLP. When infeasibility arises, we apply a correction procedure that restores feasibility while preserving the key strategic decisions as much as possible.

The proposed methods are evaluated in both a small case study and a larger case study based on Delft, the Netherlands. From an algorithmic perspective, all three methods are effective in the small case study and are able to identify good-quality solutions within reasonable computational times. In the Delft case study, the BO-based method achieves the best objective values among the three methods. These results demonstrate the potential of combining exact reformulation techniques and data-driven search methods for solving large-scale nonlinear optimisation problems in SAV operations.

From the application perspective, the three pricing strategies lead to markedly different outcomes in terms of demand, congestion, and operator profit. Pricing strategy 1 (base-fare-plus-distance-based strategy) is the most supportive of active transport for short trips, as it discourages SAV use on shorter journeys. Pricing strategy 2 (distance-based-only strategy) leads to the lowest congestion and the smallest fleet, but also produces the lowest profit. Pricing strategy 3 (income class-based pricing strategy) yields the highest profit by attracting the largest SAV demand and differentiating fares according to travellers' price sensitivity. Although this strategy results in higher congestion, it also improves accessibility for low- and middle-income users, and therefore, performs well from both an operational and equity perspective.

Finally, the additional analyses highlight the importance of accounting for income heterogeneity when designing SAV pricing strategies. Ignoring such heterogeneity can lead to biased demand estimates and unrealistic operational decisions. The results further show that pricing strategy 3 performs robustly, yielding relatively stable pricing decisions across zones with different income compositions. In addition, neglecting congestion leads to overly optimistic travel-time assumptions and, consequently, biased demand and fleet decisions. Overall, this study provides both methodological and practical insights into the design of efficient, equity-aware, and computationally tractable pricing and fleet-management strategies for SAV services.

Future research could explore several promising areas. Firstly, investigating dynamic pricing strategies that account for spatial-temporal congestion levels could provide more effective congestion management. Secondly, examining optimal ride-sharing pricing mechanisms could help alleviate congestion effects and increase fleet utilisation rates. Additionally, the pricing mechanisms of multiple SAV service providers could be studied to understand competitive dynamics and market impacts. Finally, developing heuristics to generate high-quality starting solutions for commercial solvers is a promising approach to ensure optimal outcomes.

## CRediT authorship contribution statement

**Qiaochu Fan:** Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **J. Theresia van Essen:** Conceptualization, Methodology, Supervision, Validation, Writing – review & editing; **Gonçalo Homem de Almeida Correia:** Conceptualization, Methodology, Supervision, Validation, Writing – review & editing.

## Acknowledgments

The first author acknowledges financial support from the V2G-QUESTS project (F-DUT-2022-0241), funded by the Dutch Research Council (NWO) under the Driving Urban Transitions Partnership, which is co-funded by the European Commission. The authors would like to thank the reviewers for their valuable feedback and constructive suggestions, which helped improve the quality and clarity of the manuscript.

## Data availability

Data will be made available on request.

## Appendix A. Problem formulation

See Table A.9 for the mathematical notations.

**Table A.9**
Notation.

| Notation | Description |
|---|---|
| **Set** | |
| $\mathcal{T}$ | Set of time instants $\mathcal{T} = \{0, 1, 2, \dots, T\}$ in the operational period. |
| $\mathcal{N}$ | Set of physical nodes within the network. |
| $\mathcal{L}$ | Set of road links connecting the nodes in set $\mathcal{N}$. |
| $\mathcal{G}$ | Set of links in the time-space network. |
| $\mathcal{N}_P$ | Set of nodes that allowing parking for SAVs with $\mathcal{N}_P \subseteq \mathcal{N}$. |
| $\mathcal{C}$ | Set of income classes, categorised into low income ($low$), middle income ($mid$), and high income ($high$). |
| $\mathcal{R}^c$ | Set of groups of trips associated with a specific income class $c \in \mathcal{C}$. Each group $r \in \mathcal{R}^c$ consists of trips that share the same characteristics: origin, destination, departure time, latest arrival time, and income level $c$. |
| $\mathcal{M}$ | Set of travel modes, consisting of automated vehicles ($AV$) and bicycles ($B$) as options. |
| $\mathcal{K}$ | $= \{1, 2, \dots, k, \dots, K\}$. Index set of predetermined breakpoints. |
| **Parameters** | |
| $\Delta t$ | Time step. |
| $l_{ij}$ | Length of road link $(i, j) \in \mathcal{L}$. |
| $Q_{ij}$ | Capacity of road link $(i, j) \in \mathcal{L}$ in number of vehicles per time step. |
| $t_{ij}^{\max}$ | Maximum travel time by car on road link $(i, j) \in \mathcal{L}$. |
| $t_{ij}^{\min}$ | Minimum travel time by car on road link $(i, j) \in \mathcal{L}$. |
| $C_{i_{t_1} j_{t_2}}$ | Spatial capacity of road link $(i, j) \in \mathcal{L}$ in number of vehicles that fit on the road link from time instant $t_1$ to $t_2$, where $(i_{t_1}, j_{t_2}) \in \mathcal{G}$. |
| $\alpha$ | Trip service rate when all the requests have to be accepted, %. |
| $o^r$ | Origin node for group of trips $c \in \mathcal{C}, r \in \mathcal{R}^c$. |
| $d^r$ | Destination node for group of trips $c \in \mathcal{C}, r \in \mathcal{R}^c$. |
| $a^r$ | Departure time for group of trips $c \in \mathcal{C}, r \in \mathcal{R}^c$. |
| $b^r$ | Latest arrival time for group of trips $c \in \mathcal{C}, r \in \mathcal{R}^c$. |
| $l^r_{\text{short}}$ | Shortest travel distance for group of trips $c \in \mathcal{C}, r \in \mathcal{R}^c$, in kilometres. |
| $t^r_{\text{short}}$ | Shortest travel time assuming free-flow speed for group of trips $c \in \mathcal{C}, r \in \mathcal{R}^c$, in time steps. |
| $n^r$ | Total number of trips for group $c \in \mathcal{C}, r \in \mathcal{R}^c$. |
| $V_B^r$ | Deterministic systematic component of the utility of bicycles for group of trips $c \in \mathcal{C}, r \in \mathcal{R}^c$. |
| $c_B^r$ | Monetary costs of travellers in group $c \in \mathcal{C}, r \in \mathcal{R}^c$ using bicycles, in euros. |
| $\beta_0^c$ | Parameter converting generalised costs into utility for income class $c \in \mathcal{C}$, in utility/euro. |
| $\beta_1$ | Parameter converting service rate into utility. |
| $VOT_m^c$ | Travellers' value of travel time in class $c \in \mathcal{C}$ using mode $m \in \mathcal{M}$, in euros/time step. |
| $T_B^r$ | Travel time of using bicycles for trips in group $c \in \mathcal{C}, r \in \mathcal{R}^c$. |
| $c_o$ | Unit driving operational cost of an SAV, in euros/km. |
| $c_d$ | Penalty for drop-off delay of passengers, in euros/time step. |
| $c_f$ | Depreciation cost in one hour for using an SAV, in euros/vehicle . |
| $(u^k, \ln u^k)$ | Coordinates of the $k^{\text{th}}$ breakpoint. |
| $M^r$ | Big-M parameter, where $c \in \mathcal{C}, r \in \mathcal{R}^c$. |
| $M_{ijt_1t_2}$ | Big M-parameter, where $(i, j) \in \mathcal{L}, t_1, t_2 \in \mathcal{T}$, if $t_1 < t_2 \le t_1 + t_{ij}^{\max} - t_{ij}^{\min}$. |
| **Decision variables** | |
| $V_{AV}^r$ | Deterministic systematic component of travellers' utility for using an SAV in group $c \in \mathcal{C}, r \in \mathcal{R}^c$. |
| $c_{AV}^r$ | Monetary costs of travellers in group $c \in \mathcal{C}, r \in \mathcal{R}^c$ using SAVs, in euros. |
| $T_{AV}^r$ | Longest SAVs travel time for group $c \in \mathcal{C}, r \in \mathcal{R}^c$. |
| $P_{AV}^r$ | Probability to choose SAVs for the trips in group $c \in \mathcal{C}, r \in \mathcal{R}^c$. |
| $D_{AV}^r$ | Total number of trips using SAVs in group $c \in \mathcal{C}, r \in \mathcal{R}^c$. |
| $\alpha$ | Trip service rate. |
| $p^0$ | Initial base fare for using SAVs, in euros/trip. |
| $p$ | Travel distance-related price for using an SAV, in euros/km. |
| $p^c$ | Travel distance-related price for using an SAV for income class $c \in \mathcal{C}$, in euros/km. |
| $S^r$ | Total number of trips served by SAVs from group $r$, where $c \in \mathcal{C}, r \in \mathcal{R}^c$. |
| $\hat{F}^r_{i_{t_1} j_{t_2}}$ | Passenger flow in the group of trips $c \in \mathcal{C}, r \in \mathcal{R}^c$ served by an SAV in road link $(i, j)$, from time instant $t_1$ to $t_2$. Only defined for $(i_{t_1}, j_{t_2}) \in \mathcal{G}, a^r \le t_1 < t_2 \le b^r$. If $t_1 = a^r$, then $i = o^r$. |
| $O$ | SAV fleet size. |
| $O_i$ | Initial distribution of SAVs at parking node $i \in \mathcal{N}_P$ at the beginning of a day. |
| $E_t^r$ | Total number of passengers in group of trips $c \in \mathcal{C}, r \in \mathcal{R}^c$ arriving at time $t \in \mathcal{T}$. |
| $F_{i_{t_1} j_{t_2}}$ | Vehicle flow in road link $(i, j)$ from time instant $t_1$ to $t_2$, where $(i_{t_1}, j_{t_2}) \in \mathcal{G}$. Note that when $t_1 = 0$, $i \in \mathcal{N}_P$, meaning that SAVs have to depart from the parking nodes at the beginning of a day. |
| $W_{i_t}$ | Total number of vehicles parking at node $i \in \mathcal{N}_P$ from time instant $t$ to $t + 1$, with $t \in \mathcal{T} \setminus \{T\}$. |
| $Z_t^r$ | Binary variable which is 1 when Constraint (25) is active, and 0 otherwise. |
| $X_{i_{t_1} j_{t_2}}$ | Binary variable which is 1 when any vehicle travels in road link $(i, j)$ from time instant $t_1$ to $t_2$, where $(i_{t_1}, j_{t_2}) \in \mathcal{G}$, and 0 otherwise. |
| $A_t^r$ | Binary variable which is 1 when at least one trip in group $c \in \mathcal{C}, r \in \mathcal{R}^c$ arrives at time $t \in \mathcal{T}$, and 0 otherwise. |
| $LN_{AV}^r$ | Auxiliary continuous variable, where $c \in \mathcal{C}, r \in \mathcal{R}^c$. |
| $LN_B^r$ | Auxiliary continuous variable, where $c \in \mathcal{C}, r \in \mathcal{R}^c$. |

**Table A.9**
Continued.

| Notation | Description |
|---|---|
| $\lambda_r^k$ | Binary variable indicating whether an interval $[u^k, u^{k+1}]$ is active or not, where $k \in \{1, 2, \dots, k, \dots, K-1\}$, $c \in C, r \in \mathcal{R}^c$. |
| $\theta_r^k$ | Convex combination coefficient for breakpoint $k \in K$ for group of trips $c \in C, r \in \mathcal{R}^c$. |
| $\overline{\lambda}_r^k$ | Binary variable indicating whether an interval $[1-u^{k+1}, 1-u^k]$ is active or not, where $k \in \{1, 2, \dots, k, \dots, K-1\}$, $c \in C, r \in \mathcal{R}^c$. |
| $\overline{\theta}_r^k$ | Convex combination coefficient for breakpoint $k \in K$ for group of trips $c \in C, r \in \mathcal{R}^c$. |
| $\overline{D}_h$ | Binary variables utilised for discretising integer variables, where $h \in \{0, 1, \dots, H\}$. |
| $Y_h$ | Continuous variables utilised for describing the value of the integer variables, where $h \in \{0, 1, \dots, H\}$. |
| $\overline{Y}_h^r$ | Continuous variables utilised for describing the value of the product of integer variables and continuous variables, where $c \in C, r \in \mathcal{R}^c, h \in \{0, 1, \dots, H\}$. |
| $\overline{S}_h^r$ | Binary variables utilised for discretising integer variables, where $c \in C, r \in \mathcal{R}^c, h \in \{0, 1, \dots, H\}$. |

*Mixed integer linear program*

$$\max \sum_{r\in\mathcal{R},c\in C}\sum_{h=0}^{H^r} 2^h \overline{Y}_h^r - c_f \cdot O - c_o\left(\sum_{\left(i_{t_1},j_{t_2}\right)\in\mathcal{G}} l_{ij}F_{i_{t_1}j_{t_2}}\right) - c_d \sum_{c\in C,r\in\mathcal{R}^c}\left(\sum_{t\in\mathcal{T}} tE_t^r - (a^r + t_{\text{short}}^r)S^r\right). \tag{A.1}$$

subject to:

$$c_{AV}^r = \begin{cases} p^0 + l_{\text{short}}^r p & \text{if pricing strategy 1,} \\ l_{\text{short}}^r p & \text{if pricing strategy 2,} \\ l_{\text{short}}^r p^c & \text{if pricing strategy 3,} \end{cases} \qquad \forall c \in C, r \in \mathcal{R}^c \tag{A.2}$$

$$V_{AV}^r = -\beta_0^c(c_{AV}^r + VOT_{AV}^c T_{AV}^r) - \beta_1(1-\alpha), \qquad \forall c \in C, r \in \mathcal{R}^c \tag{A.3}$$

$$n^r P_{AV}^r - 0.5 < D_{AV}^r \le n^r P_{AV}^r + 0.5, \qquad \forall r \in \mathcal{R}^c, c \in C \tag{A.4}$$

$$S^r \le D_{AV}^r, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{A.5}$$

$$S^r = \sum_{j_t \mid \left(o_{a^r}^r, j_t\right)\in\mathcal{G}} \hat{F}_{o_{a^r}^r j_t}^r, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{A.6}$$

$$S^r = \sum_{t\in\mathcal{T}\mid a^r + t_{\text{short}}^r \le t \le b^r} E_t^r, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{A.7}$$

$$E_t^r = \sum_{i_{t_1} \mid \left(i_{t_1}, d_t^r\right)\in\mathcal{G}} PF_{i_{t_1}d_t^r}^r, \qquad \forall c \in C, r \in \mathcal{R}^c, t \in \mathcal{T} \tag{A.8}$$

$$\sum_{j_{t_1} \mid \left(d_t^r, j_{t_1}\right)\in\mathcal{G}} \hat{F}_{d_t^r j_{t_1}}^r = 0, \qquad \forall c \in C, r \in \mathcal{R}^c, a^r \le t \le b^r \tag{A.9}$$

$$\sum_{i_{t_1} \mid \left(i_{t_1}, o_t^r\right)\in\mathcal{G}} \hat{F}_{i_{t_1}o_t^r}^r = 0, \qquad \forall c \in C, r \in \mathcal{R}^c, a^r \le t \le b^r \tag{A.10}$$

$$\sum_{j_{t_0} \mid \left(j_{t_0}, i_{t_1}\right)\in\mathcal{G}} \hat{F}_{j_{t_0}i_{t_1}}^r = \sum_{j_{t_2} \mid \left(i_{t_1}, j_{t_2}\right)\in\mathcal{G}} \hat{F}_{i_{t_1}j_{t_2}}^r, \qquad \forall c \in C, r \in \mathcal{R}^c, a^r < t_1 < b^r, i \in \mathcal{N}, i \ne o^r, i \ne d^r \tag{A.11}$$

$$\sum_{r\in\mathcal{R}^c c\in C} \hat{F}_{i_{t_1}j_{t_2}}^r \le F_{i_{t_1}j_{t_2}}, \qquad \forall\left(i_{t_1}, j_{t_2}\right) \in \mathcal{G} \tag{A.12}$$

$$\sum_{j_{t_1} \mid \left(j_{t_1}, i_t\right)\in\mathcal{G}, t_1 < t} F_{j_{t_1}i_t} = \sum_{j_{t_2} \mid \left(i_t, j_{t_2}\right)\in\mathcal{G}, t < t_2} F_{i_t j_{t_2}}, \qquad \forall i \in \mathcal{N} \setminus \mathcal{N}_P, 0 < t < \mathcal{T} \tag{A.13}$$

$$\sum_{j_{t_1}|\left(j_{t_1},i_t\right)\in\mathcal{G},t_1<t} F_{j_{t_1}i_t} + W_{i_{t-1}} = \sum_{j_{t_2}|\left(i_t,j_{t_2}\right)\in\mathcal{G},t<t_2} F_{i_tj_{t_2}} + W_{i_t}, \qquad \forall i \in \mathcal{N}_P, 0 < t < \mathcal{T} \tag{A.14}$$

$$\sum_{j_t|(i_0,j_t)\in\mathcal{G}} F_{i_0j_t} + W_{i_0} = O_i, \qquad \forall i \in \mathcal{N}_P \tag{A.15}$$

$$\sum_{i\in\mathcal{N}_P} O_i = O \tag{A.16}$$

$$\frac{E_t^r}{n^r} \le A_t^r \le E_t^r, \qquad \forall c \in C, r \in \mathcal{R}^c, a^r + t^r_{\text{short}} \le t \le b^r \tag{A.17}$$

$$T_{AV}^r \ge A_t^r(t - a^r), \qquad \forall c \in C, r \in \mathcal{R}^c, a^r + t^r_{\text{short}} \le t \le b^r \tag{A.18}$$

$$T_{AV}^r \le A_t^r(t - a^r) + (b^r - a^r)(1 - Z_t^r), \qquad \forall c \in C, r \in \mathcal{R}^c, a^r + t^r_{\text{short}} \le t \le b^r \tag{A.19}$$

$$\sum_{t|a^r+t^r_{\text{short}}\le t\le b^r} Z_t^r = 1, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{A.20}$$

$$\sum_{t_2|\left(i_{t_1},j_{t_2}\right)\in\mathcal{G}} X_{i_{t_1}j_{t_2}} \le 1, \qquad \forall (i,j) \in \mathcal{L}, t_1 \in \mathcal{T} \tag{A.21}$$

$$F_{i_{t_1}j_{t_2}} \le \left\lfloor C_{i_{t_1}j_{t_2}} \right\rfloor X_{i_{t_1}j_{t_2}}, \qquad \forall\left(i_{t_1}, j_{t_2}\right) \in \mathcal{G} \tag{A.22}$$

$$t_1 + \sum_{t\in\mathcal{T}} X_{i_{t_1}j_t}\left(t - t_1\right) \le t_2 + \sum_{t\in\mathcal{T}} X_{i_{t_2}j_t}(t - t_2) + (t_1 + t_{ij}^{\max} - t_2)\left(1 - \sum_{t\in\mathcal{T}} X_{i_{t_2}j_t}\right), \qquad \forall (i,j) \in \mathcal{L}, \quad t_1 < t_2 \le t_1 + t_{ij}^{\max} - t_{ij}^{\min} \tag{A.23}$$

$$S^r = \sum_{h=0}^{H^r} 2^h \overline{S}_h^r, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{A.24}$$

$$\overline{Y}_h^r \le p_{\max}\overline{S}_h^r, \qquad \forall c \in C, r \in \mathcal{R}^c, h \in \{0, 1, \ldots, H^r\} \tag{A.25}$$

$$\overline{Y}_h^r \ge c_{AV}^r - p_{\max}(1 - \overline{S}_h^r), \qquad \forall c \in C, r \in \mathcal{R}^c, h \in \{0, 1, \ldots, H^r\} \tag{A.26}$$

$$\overline{Y}_h^r \le c_{AV}^r, \qquad \forall c \in C, r \in \mathcal{R}^c, h \in \{0, 1, \ldots, H^r\}. \tag{A.27}$$

$$c_{AV}^r \le \max_{c\in C, r\in\mathcal{R}^c}\left(\frac{-V_B^r + \ln(2n^r - 1)}{\beta_0^c} - VOT_{AV}^c t^r_{\text{short}}\right), \qquad \forall c \in C, r \in \mathcal{R}^c \tag{A.28}$$

$$LN_{AV}^r - LN_B^r = V_{AV}^r - V_B^r, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{A.29}$$

$$LN_{AV}^r \le \frac{1}{u^k}P_{AV}^r + \ln u^k - 1, \qquad \forall c \in C, r \in \mathcal{R}^c, k \in K \tag{A.30}$$

$$LN_{AV}^r \ge \sum_{k=1}^{K} \theta_r^k \ln u^k, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{A.31}$$

$$P_{AV}^{r} = \sum_{k=1}^{K} \theta_{r}^{k} u^{k}, \qquad \forall c \in C, r \in \mathcal{R}^{c} \tag{A.32}$$

$$\sum_{k=1}^{K} \theta_{r}^{k} = 1, \qquad \forall c \in C, r \in \mathcal{R}^{c} \tag{A.33}$$

$$\sum_{k=1}^{K-1} \lambda_{r}^{k} = 1, \qquad \forall c \in C, r \in \mathcal{R}^{c} \tag{A.34}$$

$$\theta_{r}^{1} \le \lambda_{r}^{1}, \qquad \forall c \in C, r \in \mathcal{R}^{c} \tag{A.35}$$

$$\theta_{r}^{k} \le \lambda_{r}^{k-1} + \lambda_{r}^{k}, \qquad \forall c \in C, r \in \mathcal{R}^{c}, k \in \{2, \dots, K-1\} \tag{A.36}$$

$$\theta_{r}^{K} \le \lambda_{r}^{K-1}, \qquad \forall c \in C, r \in \mathcal{R}^{c} \tag{A.37}$$

$$LN_{B}^{r} \le \frac{1}{u^{k}}(1 - P_{AV}^{r}) + \ln u^{k} - 1, \qquad \forall c \in C, r \in \mathcal{R}^{c}, k \in \mathcal{K} \tag{A.38}$$

$$LN_{B}^{r} \ge \sum_{k=1}^{K} \overline{\theta}_{r}^{k} \ln u^{k}, \qquad \forall c \in C, r \in \mathcal{R}^{c} \tag{A.39}$$

$$1 - P_{AV}^{r} = \sum_{k=1}^{K} \overline{\theta}_{r}^{k} u^{k}, \qquad \forall c \in C, r \in \mathcal{R}^{c} \tag{A.40}$$

$$\sum_{k=1}^{K} \overline{\theta}_{r}^{k} = 1, \qquad \forall c \in C, r \in \mathcal{R}^{c} \tag{A.41}$$

$$\sum_{k=1}^{K-1} \overline{\lambda}_{r}^{k} = 1, \qquad \forall c \in C, r \in \mathcal{R}^{c} \tag{A.42}$$

$$\overline{\theta}_{r}^{1} \le \overline{\lambda}_{r}^{1}, \qquad \forall c \in C, r \in \mathcal{R}^{c} \tag{A.43}$$

$$\overline{\theta}_{r}^{k} \le \overline{\lambda}_{r}^{k-1} + \overline{\lambda}_{r}^{k}, \qquad \forall c \in C, r \in \mathcal{R}^{c}, k \in \{2, \dots, K-1\} \tag{A.44}$$

$$\overline{\theta}_{r}^{K} \le \overline{\lambda}_{r}^{K-1}, \qquad \forall c \in C, r \in \mathcal{R}^{c} \tag{A.45}$$

$$\sum_{c \in C, r \in \mathcal{R}^{c}} D_{AV}^{r} = \sum_{h=0}^{H} 2^{h} \overline{D}_{h} \tag{A.46}$$

$$Y_{h} \le \alpha, \qquad \forall h \in \{0, 1, \dots, H\} \tag{A.47}$$

$$Y_{h} \le \overline{D}_{h}, \qquad \forall h \in \{0, 1, \dots, H\} \tag{A.48}$$

$$Y_{h} \ge \alpha + \overline{D}_{h} - 1, \qquad \forall h \in \{0, 1, \dots, H\} \tag{A.49}$$

$$\sum_{h=0}^{H^{r}} 2^{h} Y_{h} = \sum_{c \in C, r \in \mathcal{R}^{c}} S^{r} \tag{A.50}$$

$$0 \leq \alpha \leq 1 \tag{A.51}$$

$$p^0 \geq 0 \tag{A.52}$$

$$p \geq 0 \tag{A.53}$$

$$p^c \geq 0, \qquad \forall c \in \mathcal{C} \tag{A.54}$$

$$\overline{D}_h \in \{0,1\}, \qquad \forall h \in \{0,1,\dots,H\} \tag{A.55}$$

$$Y_h \geq 0, \qquad \forall h \in \{0,1,\dots,H\} \tag{A.56}$$

$$\overline{S}_h^r \in \{0,1\}, \qquad \forall h \in \{0,1,\dots,H^r\} \tag{A.57}$$

$$\overline{Y}_h^r \geq 0, \qquad \forall c \in \mathcal{C}, r \in \mathcal{R}^c, h \in \{0,1,\dots,H^r\} \tag{A.58}$$

$$V_{AV}^r \geq 0, \qquad \forall c \in \mathcal{C}, r \in \mathcal{R}^c \tag{A.59}$$

$$T_{AV}^r \in \mathbb{N}^0, \qquad \forall c \in \mathcal{C}, r \in \mathcal{R}^c \tag{A.60}$$

$$P_{AV}^r \geq 0, \qquad \forall c \in \mathcal{C}, r \in \mathcal{R}^c \tag{A.61}$$

$$D_{AV}^r \in \mathbb{N}^0, \qquad \forall c \in \mathcal{C}, r \in \mathcal{R}^c \tag{A.62}$$

$$O \in \mathbb{N}^0 \tag{A.63}$$

$$O_i \in \mathbb{N}^0, \qquad \forall i \in \mathcal{N}_P \tag{A.64}$$

$$S^r \in \mathbb{N}^0, \qquad \forall c \in \mathcal{C}, r \in \mathcal{R}^c \tag{A.65}$$

$$E_t^r \in \mathbb{N}^0, \qquad \forall c \in \mathcal{C}, r \in \mathcal{R}^c, t \in \mathcal{T} \tag{A.66}$$

$$\hat{F}_{i_{t_1} j_{t_2}}^r \in \mathbb{N}^0, \qquad \forall c \in \mathcal{C}, r \in \mathcal{R}^c, (i_{t_1}, j_{t_2}) \in \mathcal{G} \tag{A.67}$$

$$F_{i_{t_1} j_{t_2}} \geq 0, \qquad \forall (i_{t_1}, j_{t_2}) \in \mathcal{G} \tag{A.68}$$

$$W_{i_t} \in \mathbb{N}^0, \qquad \forall i \in \mathcal{N}_P, t \in \mathcal{T} \tag{A.69}$$

$$Z_t^r \in \{0,1\}, \qquad \forall c \in \mathcal{C}, r \in \mathcal{R}^c, t \in \mathcal{T}, a^r + t_{\text{short}}^r \leq b^r \tag{A.70}$$

$$X_{i_{t_1} j_{t_2}} \in \{0,1\}, \qquad \forall (i_{t_1}, j_{t_2}) \in \mathcal{G} \tag{A.71}$$

$$A_t^r \in \{0,1\}, \qquad \forall c \in C, r \in \mathcal{R}^c, t \in \mathcal{T}, a^r + t^r_{\text{short}} \le t \le b^r \tag{A.72}$$

$$LN^r_{AV} \ge 0, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{A.73}$$

$$LN^r_B \ge 0, \qquad \forall c \in C, r \in \mathcal{R}^c \tag{A.74}$$

$$\theta^k_r \ge 0, \qquad \forall c \in C, r \in \mathcal{R}^c, k \in \mathcal{K} \tag{A.75}$$

$$\lambda^k_r \in \{0,1\}, \qquad \forall c \in C, r \in \mathcal{R}^c, k \in \mathcal{K} \tag{A.76}$$

$$\overline{\theta}^k_r \ge 0, \qquad \forall c \in C, r \in \mathcal{R}^c, k \in \mathcal{K} \tag{A.77}$$

$$\overline{\lambda}^k_r \in \{0,1\}, \qquad \forall c \in C, r \in \mathcal{R}^c, k \in \mathcal{K} \tag{A.78}$$

## Appendix B. Sensitivity analysis of parameters used in PSO

Parameter tuning is crucial for the performance of metaheuristic algorithms, as the selected settings can significantly influence the outcome. In this study, we conduct a sensitivity analysis on the key parameters of the PSO-based algorithm. These parameters include population size, the cognitive and social coefficients ($c_{1,\max}$ and $c_{2,\max}$), and the inertia coefficient ($w$). The scenarios with varying parameter settings are presented in Table B.10. For each scenario, we conducted five experiments and reported the mean, minimum, and maximum values of the objective functions. These results are shown in Table B.11, where only the objective function values are displayed as they serve as an indicator of model performance.

As shown in Table B.11, varying the population size from 12 (base scenario) to 8 (Scenario 1) or 16 (Scenario 2) results in very similar overall performance, with Scenario 1 and 2 showing slightly worse performance in most cases. A population size of 8 reduces diversity within the population, while a population size of 16 increases the computational time per generation, leading to fewer iterations within the same timeframe. Larger cognitive and social coefficients ($c_{1,\max}$ and $c_{2,\max}$) promote more aggressive exploration and exploitation, whereas smaller values are more conservative and help prevent overshooting optimal regions. The results indicate that the performance in the base scenario and in Scenarios 3 and 4 is quite similar, making it difficult to draw a definitive conclusion about the impact of these coefficients on the outcome.

In Scenarios 5 and 6, we varied the inertia coefficient ($w$). A higher inertia encourages particles to retain a greater portion of their previous velocity, leading to more exploratory behaviour, with particles more likely to continue moving in their current direction. Conversely, lower inertia makes particles more responsive to the cognitive and social coefficients, thereby promoting quicker convergence and increased exploitation. The results from the base scenario, Scenario 5, and Scenario 6 are very similar. In our problem setting, these parameter variations do not significantly affect the model's performance, indicating that the algorithm is able to maintain stable performance across a range of different parameter settings.

**Table B.10**
Parameter setting for sensitivity analysis.

| Scenario | Population size | $c_{1,\max}$,$c_{2,\max}$ | $w$ |
|---|---|---|---|
| Base scenario | 12 | 0.8 | 0.2 |
| Scenario 1 | 8 | 0.8 | 0.2 |
| Scenario 2 | 16 | 0.8 | 0.2 |
| Scenario 3 | 12 | 1 | 0.2 |
| Scenario 4 | 12 | 0.6 | 0.2 |
| Scenario 5 | 12 | 0.8 | 0.1 |
| Scenario 6 | 12 | 0.8 | 0.3 |

**Table B.11**
Objective function values across different scenarios under three pricing strategies.

| | Pricing strategy 1 | | | Pricing strategy 2 | | | Pricing strategy 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Min | Max | Mean | Min | Max | Mean | Min | Max |
| Base scenario | 7669.52 | 7553.52 | 7768.87 | 7161.29 | 7084.55 | 7207.06 | 8962.05 | 8778.77 | 9105.75 |
| Scenario 1 | 7547.43 | 7180.54 | 7760.51 | 7045.8 | 6931.4 | 7239.63 | 8822.34 | 8660 | 8920.54 |
| Scenario 2 | 7589.95 | 7513 | 7683.97 | 7090.33 | 6976.68 | 7241.6 | 8822.48 | 8737.16 | 8958.9 |
| Scenario 3 | 7597.93 | 7498.34 | 7790.28 | 6999.95 | 6398.17 | 7259.57 | 8814.33 | 8643.61 | 8963.8 |
| Scenario 4 | 7685.02 | 7590.01 | 7764.06 | 7158.50 | 6969.45 | 7254.77 | 8814.51 | 8659.01 | 9007.63 |
| Scenario 5 | 7654.72 | 7436.56 | 7797.37 | 6973.07 | 6423.31 | 7152.19 | 8910.99 | 8842.57 | 8974.48 |
| Scenario 6 | 7636.14 | 7549.43 | 7705.51 | 7182.61 | 7030.08 | 7289.55 | 8825.98 | 8733.11 | 8983.95 |

## References


Al-Kanj, L., Nascimento, J., Powell, W.B., 2020. Approximate dynamic programming for planning a ride-hailing system using autonomous fleets of electric vehicles. Eur. J. Oper. Res. 284 (3), 1088–1106.

Asadpour, A., Lobel, I., van Ryzin, G., 2023. Minimum earnings regulation and the stability of marketplaces. Manuf. Serv. Oper. Manag. 25 (1), 254–265.

Ashkrof, P., de Almeida Correia, G.H., Cats, O., Van Arem, B., 2022. Ride acceptance behaviour of ride-sourcing drivers. Transp. Res. C: Emerg. Technol. 142, 103783.

Bai, J., Tang, C.S., 2022. Can two competing on-demand service platforms be profitable? Int. J. Prod. Econ. 250, 108672.

Beirigo, B.A., Schulte, F., Negenborn, R.R., 2022. A learning-based optimization approach for autonomous ridesharing platforms with service-level contracts and on-demand hiring of idle vehicles. Transp. Sci. 56 (3), 677–703. Publisher: INFORMS. https://doi.org/10.1287/trsc.2021.1069

Ben-Akiva, M.E., Lerman, S.R., Lerman, S.R., et al., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. Vol. 9. MIT press.

Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization. Adv. Neural Inf. Process. Syst. 24, 2546–2554.

BicycleDutch, 2018. Dutch cycling figures. https://bicycledutch.wordpress.com/2018/01/02/dutch-cycling-figures/#:~:text=This%20is%20possibly%20because%20the,bike%20is%2013%20km%2Fh

Bösch, P.M., Becker, F., Becker, H., Axhausen, K.W., 2018. Cost-based analysis of autonomous mobility services. Transp. Policy 64, 76–91.

Brochu, E., Cora, V.M., De Freitas, N., 2010. A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv:1012.2599

Cachon, G.P., Daniels, K.M., Lobel, R., 2017. The role of surge pricing on a service platform with self-scheduling capacity. Manuf. Serv. Oper. Manag. 19 (3), 368–384.

Cadarso, L., Vaze, V., Barnhart, C., Marín, Á., 2017. Integrated airline scheduling: considering competition effects and the entry of the high speed rail. Transp. Sci. 51 (1), 132–154.

Chen, X.M., Zheng, H., Ke, J., Yang, H., 2020. Dynamic optimization strategies for on-demand ride services platform: surge pricing, commission rate, and incentives. Transp. Res. B: Methodol. 138, 23–45.

Correia, G.H., Van Arem, B., 2016. Solving the user optimum privately owned automated vehicles assignment problem (UO-POAVAP): a model to explore the impacts of self-driving vehicles on urban mobility. Transp. Res. B: Methodol. 87, 64–88.

Dafermos, S.C., Sparrow, F.T., 1969. The traffic assignment problem for a general network. J. Res. Natl. Bur. Stand. B 73 (2), 91–118.

Delft High Performance Computing Centre (DHPC), 2024. DelftBlue supercomputer (Phase 2). https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase2.

Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. Numer. Math. 1 (1), 269–271.

Dixit, M., Sivakumar, A., 2020. Capturing the impact of individual characteristics on transport accessibility and equity analysis. Transp. Res. Part D: Transp. Environ. 87, 102473.

Dong, X., Chow, J. Y.J., Waller, S.T., Rey, D., 2022. A chance-constrained dial-a-ride problem with utility-maximising demand and multiple pricing structures. Transp. Res. E: Logist. Transp. Rev. 158, 102601.

Eurofound, 2024. Developments in income inequality and the middle class in the EU. Publications Office of the European Union, Luxembourg. https://doi.org/10.2806/477653.

Fagnant, D.J., Kockelman, K.M., 2014. The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. Transp. Res. C: Emerg. Technol. 40, 1–13.

Fan, Q., van Essen, J.T., Correia, G. H.A., 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. Transp. Res. C: Emerg. Technol. 157, 104398.

Fan, Q., Van Essen, J.T., Correia, G. H.A., 2022. Heterogeneous fleet sizing for on-demand transport in mixed automated and non-automated urban areas. Transp. Res. Procedia 62, 163–170.

Fayed, L., Nilsson, G., Geroliminis, N., 2023. On the utilization of dedicated bus lanes for pooled ride-hailing services. Transp. Res. B: Methodol. 169, 29–52.

Fayed, L., Nilsson, G., Geroliminis, N., 2024. A dynamic macroscopic framework for pricing of ride-hailing services with an optional bus lane access for pool vehicles. Transp. Res. C: Emerg. Technol. 169, 104854.

Frazier, P.I., 2018. A tutorial on Bayesian optimization. arXiv:1807.02811

Gad, A.G., 2022. Particle swarm optimization algorithm and its applications: a systematic review. Arch. Comput. Methods Eng. 29 (5).

Guo, H., Chen, Y., Liu, Y., 2022. Shared autonomous vehicle management considering competition with human-driven private vehicles. Transp. Res. Part C: Emerg. Technol. 136, 103547.

Hörl, S., Becker, F., Axhausen, K.W., 2021. Simulation of price, customer behaviour and system impact for a cost-covering automated taxi system in zurich. Transp. Res. C: Emerg. Technol. 123, 102974.

Huang, K., de Almeida, C. G.H., An, K., 2018. Solving the station-based one-way carsharing network planning problem with relocations and non-linear demand. Transp. Res. C: Emerg. Technol. 90, 1–17. https://doi.org/10.1016/j.trc.2018.02.020

Huang, K., An, K., Rich, J., Ma, W., 2020. Vehicle relocation in one-way station-based electric carsharing systems: a comparative study of operator-based and user-based methods. Transp. Res. E: Logist. Transp. Rev. 142, 102081. https://doi.org/10.1016/j.tre.2020.102081

Jorge, D., Molnar, G., de Almeida Correia, G.H., 2015. Trip pricing of one-way station-based carsharing networks with zone and time of day price variations. Transp. Res. B: Methodol. 81, 461–482.

Ke, Z., Qian, S., 2023. Leveraging ride-hailing services for social good: fleet optimal routing and system optimal pricing. Transp. Res. C: Emerg. Technol. 155, 104284.

Kolarova, V., Cherchi, E., 2021. Impact of trust and travel experiences on the value of travel time savings for autonomous driving. Transp. Res. C: Emerg. Technol. 131, 103354. https://doi.org/10.1016/j.trc.2021.103354

Kolarova, V., Steck, F., Bahamonde-Birke, F.J., 2019. Assessing the effect of autonomous driving on value of travel time savings: a comparison between current and future preferences. Transp. Res. A: Policy Pract. 129, 155–169.

Li, S., Yang, H., Poolla, K., Varaiya, P., 2021. Spatial pricing in ride-sourcing markets under a congestion charge. Transp. Res. B: Methodol. 152, 18–45.

Liang, X., Correia, G.H., An, K., Van Arem, B., 2020. Automated taxis' dial-a-ride problem with ride-sharing considering congestion-based dynamic travel times. Transp. Res. C: Emerg. Technol. 112, 260–281.

Liao, F., Vleugel, J., Bösehans, G., Dissanayake, D., Thorpe, N., Bell, M., van Arem, B., de Almeida Correia, G.H., 2024. Mode substitution induced by electric mobility hubs: results from Amsterdam. Transp. Res. D: Transp. Environ. 129, 104118.
Lin, X., Zhou, Y.-W., 2019. Pricing policy selection for a platform providing vertically differentiated services with self-scheduling capacity. J. Oper. Res. Soc. 70 (7), 1203–1218.
Liu, Y., Bansal, P., Daziano, R., Samaranayake, S., et al., 2019. A framework to integrate mode choice in the design of mobility-on-demand systems. Transp. Res. C: Emerg. Technol. 105, 648–665. https://doi.org/10.1016/j.trc.2018.09.022
Lu, R., de Almeida, C. G.H., Zhao, X., Liang, X., Lv, Y., 2021. Performance of one-way carsharing systems under combined strategy of pricing and relocations. Transp. B: Transp. Dyn. 9 (1), 134–152.
Ma, M., Chen, Y., Liu, W., Waller, S.T., 2023. An economic analysis of a multi-modal transportation system with ride-sourcing services and multi-class users. Transp. Policy 140, 1–17.
Meskar, M., Aslani, S., Modarres, M., 2023. Spatio-temporal pricing algorithm for ride-hailing platforms where drivers can decline ride requests. Transp. Res. C: Emerg. Technol. 153, 104200.
Müller, C., Gönsch, J., Soppert, M., Steinhardt, C., 2023. Customer-centric dynamic pricing for free-floating vehicle sharing systems. Transp. Sci. 57 (6), 1406–1432.
Nieuwenhuijsen, M.J., Khreis, H., 2016. Car free cities: pathway to healthy urban living. Environ. Int. 94, 251–262.
Nourinejad, M., Ramezani, M., 2020. Ride-sourcing modeling and pricing in non-equilibrium two-sided markets. Transp. Res. B: Methodol. 132, 340–357.
On-Road Automated Driving (ORAD) committee, 2021. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. SAE International.
Özkan, E., 2020. Joint pricing and matching in ride-sharing systems. Eur. J. Oper. Res. 287 (3), 1149–1160.
Paneque, M.P., Bierlaire, M., Gendron, B., Azadeh, S.S., 2021. Integrating advanced discrete choice models in mixed integer linear optimization. Transp. Res. B: Methodol. 146, 26–49.
Paneque, M.P., Gendron, B., Azadeh, S.S., Bierlaire, M., 2022. A lagrangian decomposition scheme for choice-based optimization. Comput. Oper. Res. 148, 105985.
Pinto, H. K.R.F., Hyland, M.F., Mahmassani, H.S., Verbas, I.Ã., 2020. Joint design of multimodal transit networks and shared autonomous mobility fleets. Transp. Res. C: Emerg. Technol. 113, 2–20. https://doi.org/10.1016/j.trc.2019.06.010
Seshadri, R., de Palma, A., Ben-Akiva, M., 2022. Congestion tolling—dollars versus tokens: within-day dynamics. Transp. Res. C: Emerg. Technol. 143, 103836.
Spieser, K., Treleaven, K., Zhang, R., Frazzoli, E., Morton, D., Pavone, M., 2014. Toward a systematic approach to the design and evaluation of automated mobility-on-demand systems: a case study in singapore. Road Veh. Autom. , 229–245.
Steiner, K., Irnich, S., 2020. Strategic planning for integrated mobility-on-demand and urban public bus networks. Transp. Sci. 54 (6), 1616–1639.
Swersky, K., Duvenaud, D., Snoek, J., Hutter, F., Osborne, M.A., 2014. Raiders of the lost architecture: kernels for bayesian optimization in conditional parameter spaces. arXiv:1409.4011
Swersky, K., Snoek, J., Adams, R.P., 2013. Multi-task Bayesian optimization. Adv. Neural Inf. Process. Syst. 26, 2004–2012.
Taylor, T.A., 2018. On-demand service platforms. Manuf. Serv. Oper. Manag. 20 (4), 704–720.
Tong, Y., Wang, L., Zhou, Z., Chen, L., Du, B., Ye, J., 2018. Dynamic pricing in spatial crowdsourcing: a matching-based approach. In: Proceedings of the 2018 International Conference on Management of Data, pp. 773–788.
Train, K.E., 2009. Discrete Choice Methods with Simulation. Cambridge university press.
Van Arem, B., Azedeh, S.S., Snelder, M., Hoogendoorn, S., 2022. Accessibility of urban regions on a low car diet – A research agenda for digital twins. Communications in Transportation Research 2, 100077. https://doi.org/10.1016/j.commtr.2022.100077.
Van Essen, J.T., Correia, G. H.A., 2019. Exact formulation and comparison between the user optimum and system optimum solution for routing privately owned automated vehicles. IEEE Trans. Intell. Transp. Syst. 20 (12), 4567–4578.
Västberg, O.B., Karlström, A., Jonsson, D., Sundberg, M., 2020. A dynamic discrete choice activity-based travel demand model. Transp. Sci. 54 (1), 21–41.
Verbich, D., El-Geneidy, A., 2017. Public transit fare structure and social vulnerability in montreal, canada. Transp. Res. A: Policy Pract. 96, 43–53.
Wang, S., de Almeida Correia, G.H., Lin, H.X., 2022. Modeling the competition between multiple automated mobility on-demand operators: an agent-based approach. Phys. A: Stat. Mech. Appl. 605, 128033.
Wang, X., He, F., Yang, H., Gao, H.O., 2016. Pricing strategies for a taxi-hailing platform. Transp. Res. E: Logist. Transp. Rev. 93, 212–231.
Wei, K., Vaze, V., Jacquillat, A., 2022. Transit planning optimization under ride-hailing competition and traffic congestion. Transp. Sci. 56 (3), 725–749.
Williams, C. K.I., Rasmussen, C.E., 2006. Gaussian Processes for Machine Learning. Vol. 2. MIT press Cambridge, MA.
Wu, T., Zhang, M., Tian, X., Wang, S., Hua, G., 2020. Spatial differentiation and network externality in pricing mechanism of online car hailing platform. Int. J. Prod. Econ. 219, 275–283.
Zhang, K., Nie, Y.M., 2021. To pool or not to pool: equilibrium, pricing and regulation. Transp. Res. B: Methodol. 151, 59–90.
Zheng, Y., Meredith-Karam, P., Stewart, A., Kong, H., Zhao, J., 2023. Impacts of congestion pricing on ride-hailing ridership: evidence from chicago. Transp. Res. A: Policy Pract. 170, 103639.