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

Fan, Qiaochu; van Essen, J. Theresia; Correia, Gonçalo H.A.

DOI
10.1016/j.trc.2023.104398

Publication date
2023

Document Version
Final published version

Published in
Transportation Research Part C: Emerging Technologies

Transportation Research Part C 157 (2023) 104398

Contents lists available at ScienceDirect

Transportation Research Part C

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

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 ☆

Qiaochu Fan [a,*], J. Theresia van Essen [a], Gonçalo H.A. Correia [b]

[a] Delft Institute of Applied Mathematics, Delft University of Technology, 2628 CD, Delft, The Netherlands
[b] Department of Transport & Planning, Delft University of Technology, 2628 CN, Delft, The Netherlands

ARTICLE INFO

Keywords:
Fleet sizing
Shared automated vehicles
Non-linear demand
Mode choice
Traffic congestion

ABSTRACT

Shared automated vehicles (SAV) are expected to benefit the sustainable development of urban regions and alleviate the negative impacts brought by the increasing number of private cars. In this paper, we envision a future scenario where non-pooled SAVs replace private cars and provide public on-demand mobility services to satisfy the mobility needs of a city's residents. To help service providers make profitable fleet sizing and management decisions, we develop a mixed-integer non-linear programming model that considers the congestion effects and the mode choice of urban travellers in different income classes, between SAVs and bicycles. Our model optimises both strategic decisions (fleet size, initial fleet distribution, and service quality level) and operational decisions (trip assignment, vehicle routing, parking, and relocation). Travellers' preference for both transport modes is described through a binary logit model and congestion effects are described by dynamically varying travel times with respect to traffic flow in a non-linear fashion. In addition, we investigate two types of accept/reject mechanisms (mandatory vs. non-mandatory acceptance) which lead to an endogenously determined acceptance rate that can affect travellers' willingness to use SAV services. The computational challenge posed by the non-linear and non-convex nature of the model is addressed through reformulation and the use of outer-inner approximation methods combined with a breakpoint generation algorithm. We demonstrate the effectiveness of our proposed method in a case study of the city of Delft in The Netherlands, as well as a scaling analysis on three toy networks with various sizes and demand profiles. A sensitivity analysis of key parameters is carried out to assess system performance. Computational results indicate that fleet sizing decisions are influenced not only by the population's geographical distribution and land use patterns but also by the pricing strategy, unit operating costs of the SAV fleet, network congestion level, and traveller behaviour. When the price rate of using SAVs is low, the fleet sizing decisions can also be influenced by the trip accept/reject mechanism and the travellers' sensitivity to the service quality level. In addition, a low price of SAV service will attract more users but may not necessarily bring a higher profit because of the increased traffic congestion.

☆ This article belongs to the Virtual Special Issue on "Coordinated Multimodal".

* Corresponding author.
E-mail address: q.fan-1@tudelft.nl (Q. Fan).

https://doi.org/10.1016/j.trc.2023.104398
Received 2 April 2023; Received in revised form 1 September 2023; Accepted 23 October 2023
Available online 8 November 2023

1. Introduction

The idea of replacing private cars with shared mobility services and active modes of transport (walking and cycling) has gained momentum rapidly in recent years. Several main reasons are driving this shift. Firstly, the rising number of private cars has been causing stress in cities, such as the lack of parking for the existing demand, increased traffic congestion, air pollution, energy waste, and traffic accidents between cars and between cars and vulnerable road users. These effects threaten the sustainable development of urban regions. Furthermore, removing private cars within cities can lead to numerous positive impacts on public health, including the reduction of air and noise pollution, heat islands, and the occurrence of injuries (Nieuwenhuijsen and Khreis, 2016). Promoting the use of active modes of transport stands to significantly improve public health by encouraging physical activity. In addition, on-demand mobility systems like Uber and Lyft have gained popularity due to their flexible, seamless, door-to-door services. Consequently, in many cities across the world, the concept of a 'car-free city' is being considered and even adopted. Cities like Hamburg, Oslo, Helsinki, and Madrid have announced their plans to be (partly) private car-free cities, and many cities such as Bogota, Brussels, Chengdu, Copenhagen, and Paris have implemented car-free days (Nieuwenhuijsen and Khreis, 2016).

Given the promising transition towards a transport system without private cars, researchers are exploring future smart mobility solutions for urban implementation, especially considering the high costs associated with traditional public transportation provision. Among these solutions, the utilisation of shared automated vehicles (SAVs) to provide public on-demand mobility services (Liang et al., 2020; Spieser et al., 2014) stands out as one of the most promising. Various benefits have been assessed across different dimensions. As highlighted by Spieser et al. (2014), an automated mobility-on-demand (AMoD) solution has the potential to fulfil the mobility needs of the entire population with approximately one-third of the total number of private cars in operation. Moreover, Fagnant and Kockelman (2014) point out that each SAV could replace around eleven privately owned cars, which brings sizeable energy consumption and greenhouse gas emissions savings. The deployment of SAVs in the transportation system could also lead to reduced parking demand, as revealed by Zhang and Guhathakurta (2017), due to the improved intensity of vehicle utilisation and reduced usage of private vehicles. In the future mobility system, active modes of transport, such as walking and cycling, will continue to be utilised by citizens alongside the provision of public services by SAVs. Walking remains suitable for short-distance trips, while bicycles offer a competitive mode of transport for longer distances due to their numerous advantages. Bicycles are known for their flexibility, user-friendliness, sustainability, and superior environmental friendliness compared to SAVs, making them an attractive option for individuals seeking to reduce transportation costs. Thus, even with the widespread adoption of SAVs, bicycles are expected to remain prevalent in city centres.

The emergence of such a mobility system will most likely lead to a notable surge in demand for SAV services, consequently boosting the need to enhance the supply capability for on-demand responsive services across time. Simultaneously, the large-scale deployment of SAVs is anticipated to induce substantial shifts in travel behaviour and mode choice, making the future demand profile different from what we have today. Thus, estimating the underlying demand and comprehending the factors that influence the demand profile is important for SAV operators to make the right decisions in fleet sizing and management.

Demand for future SAV services and fleet management will interact with each other. Demand for SAV services has a close relationship with travellers' choice of travel mode behaviour, which is influenced by a variety of factors including price, travel time, service quality, and comfort level associated with a particular mode of transport (Ashkrof et al., 2019; Correia et al., 2019). As a mobility service provider, an SAV operator needs to manage its fleet and provide sufficient service to fulfil the mobility needs of their clients, which could be the entire population if alternative modes are restricted. Generally, decisions within an SAV operation system fall into two main categories: (1) strategic decisions determined prior to service launch (or only questioned between large periods of time), such as fleet sizing, pricing strategy, and service quality level; and (2) operational decisions made and adjusted in real-time in response to incoming requests and the dynamically evolving network status, including trip assignment, vehicle routing, parking, and relocation decisions. This demonstrates the interdependent nature of demand and supply. However, most of the existing studies regarding SAVs assume fixed and known travel demand, which is particularly unsuitable for our problem, given that the demand for the SAV service is currently quite unknown.

Another drawback of assuming travel demand as a known fixed number or as varying linearly with the service level is the oversights of travellers' response to decreased network service levels induced by traffic congestion—a factor that significantly affects demand patterns. This aspect is often disregarded in existing research on fleet sizing and management. Having in mind that the road network is highly congested (higher travel time) and/or that the travel cost is high compared with other transport modes, travellers may adjust their choices. Numerous studies have underscored the profound influence of traffic congestion on travellers' mode preferences. For instance, a preference survey conducted by Chung et al. (2012) in Cheonggyecheon stream in downtown Seoul found a 3.2% decrease in private car usage due to increased congestion, accompanied by a corresponding 3.6% increase in subway ridership. Additionally, a survey by Tennøy (2010) revealed that 33% of travellers will shift from vehicles to other modes during periods of high congestion. Therefore, congestion plays a critical role in travellers' travel decisions and should be taken into account when predicting demand for SAV services.

In the framework of operations research, most existing research on fleet management problems with traffic congestion and travellers' mode choice utilise simulation-based methods (Gurumurthy et al., 2020; Hörl et al., 2021; Oh et al., 2020; Pinto et al., 2020; Wang et al., 2022). Although simulation-based methods possess the capability to replicate intricate systems with high levels of detail, they are often time-consuming as a large number of simulations must be executed to evaluate system performance under various scenarios of fleet size and operational rules. Extensive research has been conducted on traffic assignments, aiming to comprehend how traffic congestion influences route choice and travel demand. Nevertheless, traditional traffic assignment only models the flow between origins and destinations, without considering complex planning and operational decision-making for SAV

services, such as parking location, relocation strategies, and optimal fleet size. Recent research incorporates traffic assignment into (service) network design problems to evaluate the response of travellers to the (service) network design decisions (Cai et al., 2022; Pinto et al., 2020; Xu et al., 2018b; Ye et al., 2021). Typically framed as bi-level programming models, these problems decide planning decisions at the upper level, and independently model a traffic assignment problem with mode choice at the lower level. However, incorporating the complex operational decisions of SAV services is still challenging.

The centralised control of SAVs provides an opportunity for optimising the planning and complex operational decisions through a single-level model. Unlike human drivers who often prioritise individual route preferences, SAVs can behave cooperatively by following the route guidance from the fleet operator to maximise overall profit. With a shared profit-driven aim, the planning and operational decisions can be addressed at the same level. Therefore, we propose a single-level mathematical programming model from the perspective of an SAV operator to determine the most profitable strategic decisions (fleet size, initial fleet distribution, and service quality level) alongside the operational decisions of a typical day (trip assignment, vehicle routing, parking, and relocations) while considering the congestion effects and the traveller's mode choice between SAVs and active modes of transport depending on their specific income profile. In this study, we take bicycles as the representative of the active modes of transport for the sake of simplicity. We specifically focus on commuting trips during the morning peak hour, and as such, walking is not considered a competitive mode of transportation for SAVs.

We explore two types of accept/reject mechanisms, namely (1) accepting all the requests, and (2) rejecting some requests but the rejection rate will influence travellers' attitudes towards using SAV services, to investigate how service quality levels influence the travel demand and SAV operator supply decisions. This type of accept/reject mechanism is widely considered in dial-a-ride problems when some trips are not profitable or impossible to be picked up or delivered within the desired time windows. However, most papers only consider this mechanism from the service providers' point of view and ignore the effect of service quality level on passengers' willingness to use the service again. A high rejection rate of user requests will lower the probability of a client requesting the service again.

We formulate the proposed problem as a novel mixed integer non-linear non-convex programming model, in which a binary logit model is embedded to describe the travellers' mode choice between SAVs and bikes. Recognising that travellers with different demographic characteristics behave differently in terms of mode choice, we divide the users into three income classes. Each class of travellers perceives the travel utility differently, which is time-dependent, flow-dependent, and path-dependent. The congestion effect is described within the model by dynamically varying travel times with respect to traffic flow. To capture the time-dependent features of traffic flow, we model the vehicle movement through flow variables in a time-space network where the studied time period is discretised and the spatial network is expanded in the time dimension. Different from the traditional vehicle routing problem where each vehicle is tracked individually, we use an aggregated flow which can reduce the number of decision variables in the optimisation model. To facilitate the solution process, we reformulate the model into a mixed-integer linear programming model. Linearisation techniques are proposed to tackle the non-linearity brought by the binary logit model, acceptance rate constraints, and demand calculation constraints. In particular, the outer-inner approximation method together with a breakpoint generation method is proposed to linearise the binary logit model. The breakpoint generation method aims to find the least number of breakpoints with a pre-specified acceptable approximation error. The expression of the maximum approximation error in a specified interval is given.

The main contributions of this paper are summarised as follows:

- This paper extends previous work by incorporating travellers' reactions to traffic congestion levels. The fleet management decisions in this context, including the strategic decisions and operational decisions, are constrained by the demand-supply equilibrium modelled through multi-class travellers' mode choice behaviour and are influenced by congestion effects. Thus, the obtained results can provide more practical and realistic managerial and operation insights for future SAV operators.
- This paper investigates how accept/reject mechanisms influence the closed demand-supply loop in the context of SAVs. It also highlights the importance of considering the impact of service quality on passengers' willingness to continue using SAV service, emphasising the role of user attitudes and satisfaction in the success and sustainability of SAV systems. To the best of our knowledge, it is still an under-explored topic.
- The proposed flow-based model incorporates three essential elements – mode choice, traffic congestion, and accept/rejection mechanisms – into a single-level optimisation model. This allows us to explore the intricate interactions between these elements, leading to a better understanding of the dynamics in SAV systems.
- The proposed model is validated through comprehensive case studies conducted in the city of Delft, The Netherlands, and three toy networks with various sizes and demand profiles. Additionally, sensitivity analyses on critical parameters are conducted, enabling a thorough assessment of system performance under diverse scenarios.

The remaining sections of this paper are organised as follows. The literature on fleet management and demand modelling is reviewed in Section 2. Section 3 presents the non-linear non-convex mathematical model of the proposed fleet sizing and management problem. In Section 4, a detailed description is provided of how to linearise the proposed model, enabling its solvability using state-of-the-art solvers. In Section 5, a case study on the city of Delft in the Netherlands is performed. In addition, a scaling analysis is conducted in Section 6 to evaluate the model's performance with various network sizes and demand profiles. Section 7 gives the main conclusions and provides an outlook on research needs.

2. Literature review

In this paper, we aim to combine the SAV service fleet sizing and management problem, SAVs congestion modelling, and SAV demand modelling in one optimisation problem. Therefore, from these three aspects, we review the literature to demonstrate the gaps that we have identified as well as search the grounds for the required methodologies for our purpose.

2.1. Fleet management with congestion effect and travellers' mode choice

A considerable amount of literature has been published on the fleet management problem. The literature includes a wide range of topics such as capacitated vehicle routing problems, vehicle routing problems with time windows, pickup and/or delivery problems, fleet sizing and vehicle routing problems, dial-a-ride transport, etc. This is independent of the vehicles being or not automated. Interested readers can refer to Hyland and Mahmassani (2017) for the taxonomy of vehicle fleet management problems. More specifically, we classify our problem as an extension of fleet sizing and the vehicle routing problem with travellers' pickup and delivery. The objective of this type of problem is to identify the optimal planning decisions that yield the minimum costs or the maximum profit for the fleet operator, which is constrained by the trip assignment and vehicle operations on the network.

As we mentioned before, most of the existing research on fleet management problems with closed supply–demand loops, the consideration of congestion effect and travellers' mode choice use simulation techniques (Gurumurthy et al., 2020; Hörl et al., 2021; Oh et al., 2020; Pinto et al., 2020; Wang et al., 2022). Only a handful of studies consider a similar problem using an optimisation-based method or a hybrid approach that combines optimisation with simulation. Wei et al. (2022) study the optimal transit schedules while taking into account the competition with ride-hailing services and traffic congestion. A mixed integer non-linear program is proposed and solved using a bi-level heuristic algorithm including an outer loop and inner loop. The strategic transit scheduling decisions are determined in the outer loop given travellers' mode choice and congestion estimates. The path choice of ride-hailing vehicles and congestion levels are determined in the inner loop in a traffic assignment problem. However, they consider simplified ride-hailing operations by ignoring the relocation of ride-hailing vehicles, and parking decisions. Thus, the congestion effects caused by the re-locations of vehicles cannot be captured in the model. Pinto et al. (2020) combine optimisation-based and simulation-based techniques. They propose a bi-level programming model to investigate the integration of the transit network redesign problem and fleet sizing problem for a shared autonomous mobility service. The upper level determines the transit pattern headways and fleet size of SAVs and the lower level describes the combined mode choice–traveller assignment problem. Their approach involves an iterative heuristic procedure where the upper-level problem is solved with a non-linear programming solver and the lower-level problem is solved through agent-based simulation given the decisions made at the upper level.

2.2. Congestion modelling in fleet management problems

Congestion in road transportation networks has been extensively studied in traffic assignment problems. As one of the major factors influencing the transportation network's performance and decisions related to route choice, congestion effects are increasingly being considered in fleet management problems as well. Liang et al. (2018) envision a future on-demand mobility system where automated vehicles (AVs) serve as taxis to provide mobility services. They take into account the impact of congestion in determining the optimal trip assignment and dynamic routing of AVs. Expanding on this theme, Liang et al. (2020) delve into a dial-a-ride problem involving ride-sharing in light of the traffic congestion caused by the routing of a large number of AVs. Fan et al. (2022) investigate the heterogeneous fleet sizing and vehicle routing problem for an on-demand mobility service provider envisioning a progressive expansion of AVs-only zones. The congestion effect is incorporated into the model to examine the impacts of the AVs-only zone on travellers' behaviour and network performance.

The congestion effect, measured quantitatively by the variation of travel time as a function of flow, is typically expressed by the well-known Bureau of Public Roads (BPR) function, which is non-linear. Incorporating this non-linear function into a mathematical programming model makes it difficult to solve to optimality. To address this issue, techniques have been proposed including but not limited to: (1) reformulating and linearising the non-linear term (Wang et al., 2015); (2) replacing the BPR function by selecting one from multiple link-traveltime choices at each time point (Van Essen and Correia, 2019); (3) adopting an iterative solution algorithm until the algorithm converges (Correia and Van Arem, 2016). In addition to the mathematical programming model, simulation-based methods have also been used as a modelling technique to study the congestion effects on the fleet management problem (Fagnant and Kockelman, 2014; Wang et al., 2022).

Existing studies on SAV fleet management problems fall short of taking users' preferences and choice behaviours into account, especially the behaviour that is influenced by the effect of traffic congestion on travel times. This is an important factor that impacts the mobility pattern and mode preference of travellers, thereby influencing the demand for SAV services and the supply.

2.3. Demand modelling methods in optimisation

A fundamental assumption used in a large body of literature on fleet management is that demand for all OD pairs is fixed and known in advance (Correia and Van Arem, 2016; Liang et al., 2018; Van Essen and Correia, 2019; Liang et al., 2020; Fan et al., 2022),

which does not match with the real world. Assuming travel demand to be constant may lead to unrealistic managerial decisions that result in substantial financial losses for SAV operators. Thus, a more appropriate representation of demand in the fleet management problem is essential.

Demand modelling methods have been widely explored in the existing literature evolving from trip-based models to activity-based models. Trip-based demand modelling representations include but are not limited to the following: (a) elastic demand represented by a simple linear function (Jorge et al., 2015) or a non-linear function such as an exponential function (Huang and Kockelman, 2020; Huang et al., 2020; Xu and Meng, 2020); (b) probability-based demand representation of discrete choice models, such as the binary logit model (Guo et al., 2022; Lu et al., 2021; Tian et al., 2022), multinomial logit model (Atasoy et al., 2014; Joksimovic et al., 2005; Yang et al., 2022), logit-based chance-constrained model (Dong et al., 2022), mixed logit model (You et al., 2022); (c) disaggregate demand representation of discrete choice models by using simulation-based linearisation (Paneque et al., 2021, 2022); (d) machine-learning methods under a big data context (Wang et al., 2020). In addition to these trip-based methods, researchers have also developed methods to study activity-based models that focus on the interdependent choice of full daily activity-travel patterns at an individual or household level. These include nested logit models, dynamic discrete choice models (Västberg et al., 2020), machine-learning-based methods (Ren and Chow, 2022), etc.

Among these methods, the discrete choice model is traditionally used to analyse choice behaviours. In this paper, we incorporate the logit model into our trip-based optimisation problem for the following reasons. Compared with a simple linear function or an exponential function, the logit model is more realistic as it describes the probability of selecting a particular alternative against other alternatives considering a number of factors and their relative importance. It can also take into account the travellers' socioeconomic characteristics such as income level, age, etc. Simulation-based linearisation method is a promising method, but generating a large number of scenarios may bring a big computational burden. Machine learning methods can analyse individual decisions with a higher prediction accuracy but the big data context is missing for the future scenario.

Including the non-linear logit formula in a mathematical model makes the model difficult to solve to optimality due to the non-linear and non-convex formulations. To tackle the non-linearity, researchers proposed many solution methods, such as linearisation algorithm (piece-wise linear function approximation(Wang and Lo, 2010), outer-approximation(Xu et al., 2018a), outer-inner approximation(Liu and Wang, 2015; Guo et al., 2022)), heuristic and meta-heuristic (Azadeh et al., 2022; Dong et al., 2022; Joksimovic et al., 2005; Tian et al., 2022; Lu et al., 2021), and simulation (Lou et al., 2011; Paneque et al., 2021; Wang et al., 2022). Among all of them, one of the most fundamental methods is the piecewise-linear function-based approximation, which aims to find the optimal solution by replacing the non-linear term in the objective function or constraints with a series of piece-wise linear functions. The key idea is to transform the mixed-integer nonlinear programming model into a linear one, and then solve the problem to optimality. A variant of this is the outer/outer-inner approximation method. Instead of replacing the non-linear term with a series of piecewise linear functions, this method specifies the upper bound/the upper and lower bound of the non-linear term using a series of linear constraints. In this paper, we adopt the outer-inner approximation method to tackle the computational challenge brought by the logit model.

3. Problem formulation

In this section, we start by outlining the assumptions of the problem and then provide a detailed description of the mathematical formulation of the proposed model.

3.1. Assumptions

In this paper, we envision a future scenario in which cities are only accessible by SAV services and active modes of transportation (e.g. bicycles). Travellers who choose SAVs can request transportation services at any location in a city using SAV service applications on their smartphones by providing trip information. After receiving the trip information, the platform decides whether to accept or reject the trip. Two accept/reject mechanisms are investigated in this paper, namely (1) an SAV operator has to accept all the requests, or (2) an SAV operator may reject a trip if it provides no benefits to the company. In this case, those rejected trips will use bicycles. Of course, travellers can choose to use bicycles directly if they perceive that using this mode provides greater travel utility. Once the request is accepted, the platform will match available SAVs with customers and dispatch them to pick up the customers.

Before we can formally introduce the model, we describe the made assumptions.

(a) The proposed model serves a strategic planning purpose. During the study period, the total mobility demand in an urban area is assumed to be constant and known in advance, enabling the SAV operator to make optimal planning decisions. This stands in contrast to real-time SAV operating systems where future demand remains uncertain, necessitating the continuous updating of the optimal operational strategy in response to new incoming demand.

(b) We assume that SAVs and bicycles operate in separate lanes or designated areas, ensuring physical separation and minimal interaction. The flexibility and manoeuvrability of bicycles allow cyclists to easily bypass congested areas and find alternative routes. As a result, our study primarily focuses on analysing the congestion effects caused by the routing of SAVs. It is worth noting the importance of considering interactions between bicycles and AVs, especially in cases where infrastructures do not allow for the separation of traffic (Hulse, 2023; Madigan et al., 2019; Vlakveld et al., 2020). However, we do not extensively delve into this topic within the scope of our paper.

(c) We assume that travellers have perfect information about the transportation network status and will make a rational mode choice based on their perceptions of travel time.

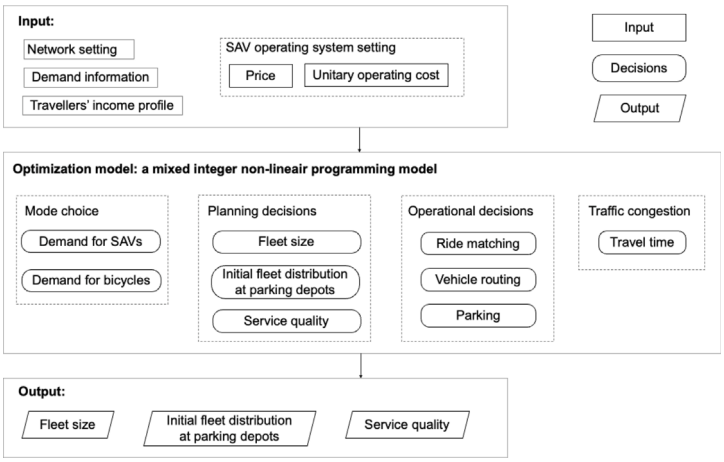


Fig. 1. Model structure.

(d) We assume that the SAVs in our study are at SAE level 5, and they are capable of driving throughout the entire network without the presence of a human driver.

(e) Neither privately-owned vehicles nor human-driven vehicles are considered as an option.

(f) A traveller will only utilise a single travel mode. Transferring between modes is not considered.

(g) The model considers exogenous fares by taking SAVs. Deciding on the optimal fares is an interesting future direction.

(h) Pooled services are not considered in this study.

To provide a comprehensive understanding of the proposed model, we have illustrated its structure in Fig. 1, which outlines the decisions, input, and output information. Additionally, we have summarised the mathematical notations used in Section 3 in Table 1 for easy reference.

3.2. Network representation

To capture the endogenous dynamic traffic congestion caused by the large-scale deployment of SAVs, we utilise a time-space network. A time-space network is a time-expanded version of the directed physical network (N, L), where N and L denote the set of nodes and road links, respectively. The time dimension is discretised into $\mathcal{T}$ periods, with each period having a duration Δt, referred to as the time step, hereinafter. Consequently, an index set of time periods $T = \{0, 1, 2, \ldots, \mathcal{T}\}$ is defined within the study horizon. At each time instant $t \in T$, the network is replicated, thus multiple networks are defined along with the time period, as shown in Fig. 2. We define set G to denote the set of links in the time-space network.

Different from the traditional physical network, the status of vehicles on the time-space network is described by both the action of the vehicles and the time those actions take. Thus, vehicles either move with passengers or relocate without passengers on links $(i_{t_1}, j_{t_2}) \in G$, representing the flow departing from node $i \in N$ to node $j \in N$ from time instant $t_1 \in T$ to time instant $t_2 \in T$, or park at node $i \in N_p$ from time instant t to $t+1$ where $t \in T$. N_p denotes the subset of nodes N allowing parking for SAVs. Here, the parking depots are restricted parking areas spread across the city that are provided by the SAV operator for their vehicles; therefore, parking is only permitted at certain designated nodes. Fig. 2 provides an illustrative example of vehicle movements within a time-space network, depicting passenger deliveries using solid lines, relocation movements without passengers heading to parking stations or passengers' pick-up/drop-off locations with dashed lines, and parking states displayed by dotted lines. In this small physical network, we assume a uniform travel time of 1 time step for each link.

Several physical attributes related to the road links are defined, such as the length of road link $(i, j) \in L$ denoted as l_{ij}, the capacity of road link $(i, j) \in L$ per time step denoted as Q_{ij}, and the maximum and minimum travel time of link $(i, j) \in L$ denoted as $t_{ij}^{\max}$ and $t_{ij}^{\min}$, respectively. Given the maximum and minimum travel time of link $(i, j) \in L$, we can further shrink the size of set G by only including the possible time choices instead of doing a complete enumeration for all the time instants. It means that we only include link (i_{t_1}, j_{t_2}) where $t_1 + t_{ij}^{\min} \le t_2 \le t_1 + t_{ij}^{\max}$.

Please note that when using a time-space network framework, travel time is represented in an integer number of time periods. However, this representation does not imply that the actual travel time must be an integer value. The integer value corresponds to the index of the time period within the study horizon. The actual value of travel time depends on the chosen time step, which may or may not be an integer. However, we do recognise that the time step sets the precision of travel times where the maximum precision is always limited by the duration of the time step.

Table 1
Notation.

Notation	Description
Sets	
T	$= \{0, 1, 2, \ldots, \mathcal{T}\}$. Set of time instants in the operation period.
N	Set of nodes.
L	Set of road links between nodes in set N.
G	Set of links in the time-space network.
N_P	Set of nodes allowing parking for SAVs with $N_P \subseteq N$.
R	Set of groups of trips, where each group of trips $r \in R$ has the same origin, destination, departure time, and latest arrival time at the destination.
M	Set of travel modes, with the automated vehicles (AV) and bicycles (B) as the two options.
Choice model	
Parameters	
V_B^r	Deterministic systematic component of the utility of bicycles for group of trips $r \in R$.
OM_m^r	Monetary costs of travellers in group $r \in R$ using mode $m \in M$, in euros.
β_0	Parameter converting generalised costs into utility, in utility/euro.
β_1	Parameter converting service rate into utility.
VOT_m^r	Travellers' value of travel time in group r using mode $m \in M$, euros/time step.
T_B^r	Travel time of using bicycles for trips in group $r \in R$.
n^r	Total number of trips for group $r \in R$.
Auxiliary variables	
V_{AV}^r	Deterministic systematic component of travellers' utility for using an SAV in group $r \in R$.
T_{AV}^r	Longest SAVs travel time for group $r \in R$.
P_{AV}^r	Probability to choose SAVs for the trips in group $r \in R$.
D_{AV}^r	Total number of trips using SAVs in group $r \in R$.
Fleet sizing and management model	
Parameters	
Δt	Time step.
l_{ij}	Length of road link $(i, j) \in L$.
Q_{ij}	Capacity of road link $(i, j) \in L$ in vehicles per time step.
$t_{ij}^{\max}$	Maximum travel time by cars on road link $(i, j) \in L$.
$t_{ij}^{\min}$	Minimum travel time by cars on road link $(i, j) \in L$.
$C_{i_{t_1} j_{t_2}}$	Spatial capacity of road link $(i, j) \in L$ in vehicles that fit on the road link from time instant t_1 to t_2, where $(i_{t_1}, j_{t_2}) \in G$.
α	Trip service rate when all the requests have to be accepted, %.
o^r	Origin node for group of trips $r \in R$.
d^r	Destination node for group of trips $r \in R$.
a^r	Departure time for group of trips $r \in R$.
b^r	Latest arrival time for group of trips $r \in R$.
sd^r	Shortest travel distance for group of trips $r \in R$, in kilometres.
st^r	Shortest travel time assuming free-flow speed for group of trips $r \in R$, in time steps.
p^0	Initial base fare for using an SAV, in euros/trip.
p	Travel distance-related price for using an SAV, in euros/km.
co	Unit driving operational cost of an SAV, in euros/km.
cd	Penalty for drop-off delay of passengers, in euros/time step.
cf	Depreciation cost in one hour for using an SAV, in euros/vehicle.
Decision variables	
S^r	Total number of trips served by SAVs from group r, where $r \in R$.
$PF_{i_{t_1} j_{t_2}}^r$	Passenger flow in the group of trips $r \in R$ served by an SAV in road link (i, j), from time instant t_1 to t_2. Only defined for $(i_{t_1}, j_{t_2}) \in G, a^r \le t_1 < t_2 \le b^r$. If $t_1 = a^r$, then $i = o^r$.
Auxiliary variables	
α	Trip service rate when some requests can be rejected.
V	SAV fleet size.
V_i	Initial distribution of SAVs at parking node $i \in N_p$ at the beginning of a day.
E_t^r	Total number of passengers in group of trips $r \in R$ arriving at time $t \in T$.
$F_{i_{t_1} j_{t_2}}$	Vehicle flow in road link (i, j) from time instant t_1 to t_2, where $(i_{t_1}, j_{t_2}) \in G$. Note that when $t_1 = 0$, $i \in N_P$, 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 $i \in N_P$ from time instant t to $t + 1$, with $t \in T$.
Z_t^r	Binary variable with $r \in R, t \in T$ if $a^r + st^r \le b^r$.
$X_{i_{t_1} j_{t_2}}$	Binary variable which is 1 when any vehicle travels in road link (i, j) from time instant t_1 to t_2, where $(i_{t_1}, j_{t_2}) \in G$, and 0 otherwise.
A_t^r	Binary variable which is 1 when at least one trip in group $r \in R$ arrives at time $t \in T$, and 0 otherwise.

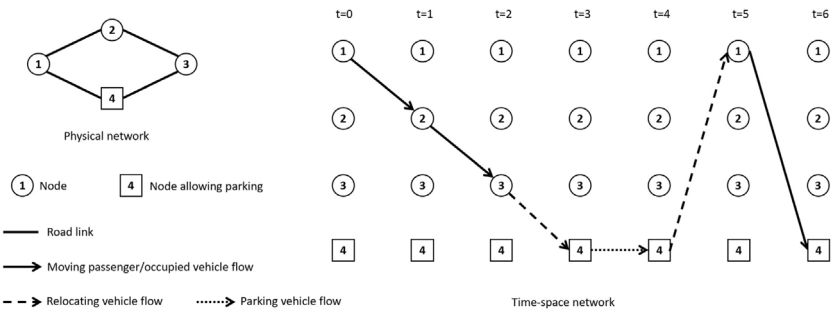


Fig. 2. Illustration of the physical network, the time-space network, and vehicle movements over the time-space network.

3.3. Demand representation and choice modelling

We consider the peak hour of a typical workday in an urban area, where congestion occurs which influences travellers' mode choices. Instead of tracking each trip individually, demand that shares the same travel information is aggregated into a group. We introduce set R as the set of groups of trips, where each group of trips $r \in R$ has the same origin $o^r \in N$, destination $d^r \in N$, departure time $a^r \in T$, latest arrival time $b^r \in T$, shortest travel distance sd^r in kilometres, shortest travel time st^r, and the same income level in euro/time step. Note that trips within a group only have the same departure time, latest arrival time, and shortest travel time based on time periods, not based on the real times in seconds. Trips from one group are allowed to use different transport modes and the total number of trips in group $r \in R$ is denoted as n^r. We denote M as the set of travel modes, with the shared automated vehicles (AV) and bicycles (B) as the two options.

The mode choices for travellers in each group are analysed using discrete choice modelling within the framework of random utility maximisation theory. To measure the willingness or preference for using a certain type of travel mode, the utility of each mode is calculated. Besides that, three income levels for travellers are considered: high income, middle income, and low income. The income level determines the value of travel time (VOTT) of travellers, which has a direct impact on their perceived utility for using a particular travel mode and consequently influences their mode choice. The VOTT of using travel mode $m \in M$ for travellers in the same group $r \in R$ is assumed to be the same, which is denoted by VOT_m^r in euro/time step. The three classes are not explicitly defined in the model but are implicitly included in the travel information of each group of trips.

However, the utility is not known with certainty due to factors such as unobserved variation among travellers, unobserved attributes of the alternatives, and perception errors of travellers (Ben-Akiva et al., 1985). Therefore, the utility of mode $m \in M$ for trips in the group $r \in R$, denoted as U_m^r, is treated as a random variable. It consists of a deterministic systematic component V_m^r, which is the observable utility of mode $m \in M$ for trips in the group $r \in R$, and a random component ϵ_m^r, which is the unobservable component of the utility.

$$U_m^r = V_m^r + \epsilon_m^r, \qquad \forall r \in R, m \in M \tag{1}$$

The deterministic term V_{AV}^r of the utility for using an SAV for group $r \in R$ depends on the generalised cost of using SAVs and travellers' satisfaction towards SAV services. To be more specific, the generalised cost of using SAVs for group of trips $r \in R$ is calculated as the linear sum of the fare of travellers OM_{AV}^r and the travel time-related cost of the journey $VOT_{AV}^r T_{AV}^r$. The fare OM_{AV}^r of using the SAV service for travellers in group $r \in R$ depends on the pricing strategy of SAV operators and the travel distance of a trip. The perceived travel time T_{AV}^r is affected by the dynamically varying traffic congestion, which can be determined endogenously by solving the optimisation model. β_0 is a parameter that converts generalised costs into utility, expressed as utility/euro, which indicates the sensitivity of travellers to the change in the monetary costs. Travellers' satisfaction with the SAV service depends on the trip service rate α. When the SAV operator is not allowed to reject any trips, the trip service rate equals 1 (100%). However, if some trips are rejected by the SAV operator, the trip service rate will be less than 1, which will decrease the traveller's satisfaction with the SAV service. β_1 is the parameter that describes travellers' satisfaction with the service rate.

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

Alternatively, the deterministic term V_B^r of the utility for using a bicycle for group $r \in R$ is calculated based on the monetary cost of using bicycles, as shown in Eqs. (3). The monetary cost OM_B^r for group $r \in R$ is the bicycle's depreciation cost which is calculated by dividing the bicycle's purchase price by its service life. We assume that the travel time T_B^r for group $r \in R$ is a constant, as the congestion in motor lanes will not affect the travel time of bicycles.

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

In Eqs. (1), ϵ_m^r is the error between the actual utility and the systematic utility of mode $m \in M$ for trips in group $r \in R$, which can be viewed as the part of utility that is unknown to the analyst. Assuming these error terms are all independently and identically

Gumbel distributed, we can compute the probability of choosing SAVs against bicycles in group $r \in R$, denoted as continuous variables P^r_{AV}, by using a binary logit model shown in Eqs. (4).

$$P^r_{AV} = \frac{e^{V^r_{AV}}}{e^{V^r_{AV}} + e^{V^r_B}}, \qquad \forall r \in R \tag{4}$$

We introduce integer variables D^r_{AV} to represent the demand for SAVs for group $r \in R$, which can be calculated using the total number of trips n^r in group $r \in R$ multiplied by their probability of choosing SAVs. Then, we round this value to the nearest integer using a floor function as shown in Eqs. (5).

$$D^r_{AV} = \lfloor n^r P^r_{AV} + 0.5 \rfloor, \qquad \forall r \in R \tag{5}$$

3.4. Fleet sizing and management for SAV operators

In this section, we develop the base formulation for an SAV operator to manage the SAV fleet. Three tiers of decisions are made in this model: (1) at the strategic level: the overall SAV fleet size, the initial fleet distribution at the beginning of a day and the service quality level; (2) at the operational level: the assignment of passengers to SAVs, vehicle routes determination, parking and relocation decisions; and (3) the travel time on each road link.

For each group $r \in R$, the total number of trips served by SAVs is specified by integer variables S^r. Therefore, the relationship between the total number of served trips, the total demand and the service rate can be described by Constraint (6). It should be noted that when an SAV operator must accept all the requests to maintain a high level of service quality, the parameter α is set to 1. When an SAV operator can reject those requests that bring no profits for the company, α is defined as a continuous variable where $\alpha \in [0, 1]$ and its value is determined endogenously by solving the model. As a result, Constraint (6) becomes a non-linear constraint.

$$\alpha \sum_{r \in R} D^r_{AV} = \sum_{r \in R} S^r \tag{6}$$

In addition, for each of the group $r \in R$, the number of served trips S^r should be less than the demand for SAVs D^r_{AV}.

$$S^r \le D^r_{AV}, \qquad \forall r \in R \tag{7}$$

The movement of vehicles is modelled as flow circulating on the time-space network. We introduce integer variables $PF^r_{i_{t_1} j_{t_2}}$ to represent the passenger flow in the group of trips $r \in R$ served by an SAV in road link (i, j), from time instant t_1 to t_2. These variables are only defined for $(i_{t_1}, j_{t_2}) \in G, a^r \le t_1 < t_2 \le b^r$. If $t_1 = a^r$, then $i = o^r$. Passengers in the same group $r \in R$ are picked up by the SAVs at the origin node o^r at the departure time a^r where $r \in R$. This is ensured by Constraints (8). In this study, we do not model the explicit waiting time from the passenger's perspective. We focus on a strategic planning problem, assuming that all the trip information is available in advance. This enables the service operator to make optimal planning decisions. Our primary focus is to determine the required number of vehicles to ensure that travellers depart at their desired times. In addition, this analysis concentrates on two urban travel modes: SAVs and bicycles. Our expectation is for the SAV operator to deliver a high-quality service, with a short waiting time for travellers before pick-up. Nevertheless, we do acknowledge that our approach possesses limitations in scenarios where passengers make their requests without sufficient advance notice and demand immediate vehicle availability. In such cases, ignoring passenger waiting time could impact strategic decision-making. However, the explicit inclusion of waiting time through modelling could increase the model's complexity, presenting challenges in achieving optimal solutions. As a result, to strike the balance between computational complexity and accuracy, we choose not to model the waiting time explicitly.

$$S^r = \sum_{j_t \mid \left(o^r_{a^r}, j_t\right) \in G} PF^r_{o^r_{a^r} j_t}, \qquad \forall r \in R \tag{8}$$

When delivering passengers to their destination d^r, SAVs serving the same group of trips $r \in R$ are allowed to take alternative routes to evenly distribute the flow and alleviate the network burden. For this reason, SAVs may arrive at the destination at different times, but not later than the user-specified latest arrival time b^r. To describe this, integer variables E^r_t are defined to represent the total number of passengers in group of trips $r \in R$ arriving at time $t \in T$ with $a^r + st^r \le t \le b^r$. Constraints (9) and (10) ensure that the number of served trips in group $r \in R$ is equal to the number of trips arriving at destination d^r at different times.

$$S^r = \sum_{t \in T \mid a^r + st^r \le t \le b^r} E^r_t, \qquad \forall r \in R \tag{9}$$

$$E^r_t = \sum_{i_{t_1} \mid \left(i_{t_1}, d^r_t\right) \in G} PF^r_{i_{t_1} d^r_t}, \qquad \forall r \in R, a^r + st^r \le t \le b^r \tag{10}$$

Constraints (11) and (12) guarantee that the destination node d^r and the origin node o^r, respectively, of the group of trips $r \in R$ will only be visited once during client delivery. It indicates that passengers will be dropped off at the destination node the first time the SAV arrives there, and SAVs will not return to the origin node after departure.

$$\sum_{j_{t_1}|\left(d_t^r, j_{t_1}\right)\in G} PF^r_{d_t^r j_{t_1}} = 0, \qquad \forall r \in R, a^r \le t \le b^r \tag{11}$$

$$\sum_{i_{t_1}|\left(i_{t_1}, o_t^r\right)\in G} PF^r_{i_{t_1} o_t^r} = 0, \qquad \forall r \in R, a^r \le t \le b^r \tag{12}$$

The next constraints describe the passenger flow conservation at any nodes $i \in N$ in the network except the origin node o^r and destination node d^r for the group of trips $r \in R$.

$$\sum_{j_{t_1}|\left(j_{t_1}, i_t\right)\in G} PF^r_{j_{t_1} i_t} = \sum_{j_{t_2}|\left(i_t, j_{t_2}\right)\in G} PF^r_{i_t j_{t_2}}, \qquad \forall r \in R, a^r < t < b^r, i \in N, i \ne o^r, i \ne d^r \tag{13}$$

On the road, SAVs have three statuses: (1) transporting a passenger; (2) driving empty to pick up the next passenger or driving to a parking depot; and (3) being parked at a depot. We introduce continuous variables $F_{i_{t_1} j_{t_2}}$ to describe the vehicle flow (the number of SAVs) in road link (i, j) from time instant t_1 to t_2, where $(i_{t_1}, j_{t_2}) \in G$. Note that when $t_1 = 0$, only links with $i \in N_P$ are defined, meaning that SAVs have to depart from the parking nodes at the beginning of the day. In addition, continuous variables W_{i_t} are defined to represent the total number of SAVs parking at node $i \in N_P$ from time instant t to $t+1$, with $t \in T$. Note that there is no need to define $F_{i_{t_1} j_{t_2}}$ and W_{i_t} as integer variables explicitly. The integrality requirement for these variables is implicitly satisfied through Constraints (14)–(17), which will be explained later.

No matter in which status, the vehicle flows $F_{i_{t_1} j_{t_2}}$ in the time-space network should always be greater than the passenger flow $\sum_{r\in R} PF^r_{i_{t_1} j_{t_2}}$ to satisfy the mobility need, as indicated by Constraints (14).

$$\sum_{r\in R} PF^r_{i_{t_1} j_{t_2}} \le F_{i_{t_1} j_{t_2}}, \qquad \forall \left(i_{t_1}, j_{t_2}\right) \in G \tag{14}$$

The next constraints describe the flow conservation rules applied to SAVs' circulation at both normal and parking nodes.

$$\sum_{j_{t_1}|\left(j_{t_1}, i_t\right)\in G, t_1 < t} F_{j_{t_1} i_t} = \sum_{j_{t_2}|\left(i_t, j_{t_2}\right)\in G, t < t_2} F_{i_t j_{t_2}}, \qquad \forall i \in N \setminus N_P, 0 < t < \mathcal{T} \tag{15}$$

$$\sum_{j_{t_1}|\left(j_{t_1}, i_t\right)\in G, t_1 < t} F_{j_{t_1} i_t} + W_{i_{t-1}} = \sum_{j_{t_2}|\left(i_t, j_{t_2}\right)\in G, t < t_2} F_{i_t j_{t_2}} + W_{i_t}, \qquad \forall i \in N_P, 0 < t < \mathcal{T} \tag{16}$$

It is worth mentioning that the vehicle flow $F_{i_{t_1} j_{t_2}}$ is associated with a vehicle flow-related cost in the objective function, which is minimised. Further details about the objective function can be found in Section 3.6. Consequently, the minimum number of vehicles required to transport passengers in link (i_{t_1}, j_{t_2}) equals the total number of passengers $\sum_{r\in R} PF^r_{i_{t_1} j_{t_2}}$ according to Constraints (14), which is an integer. Besides the occupied vehicle flow, $F_{i_{t_1} j_{t_2}}$ also includes the empty relocating vehicle flow. These empty vehicles are either driving after delivering the passengers or are on their way to pick up new passengers. According to the vehicle conservation constraints, these flows may be integer values as well.

The initial distribution of SAVs at parking node $i \in N_p$ at the beginning of a day is denoted by integer variables V_i. At the beginning of the optimisation period, SAVs either depart from the parking depots to pick up passengers or park at the parking node waiting for the task given by the SAV operator, as described in Constraints (17). Given that both $F_{i_0 j_t}$ and V_i take integer values, W_{i_0} will also take integer values.

$$\sum_{j_t|(i_0, j_t)\in G} F_{i_0 j_t} + W_{i_0} = V_i, \qquad \forall i \in N_p \tag{17}$$

In addition, the sum of the initial fleet distributed at the parking nodes gives the overall SAV fleet size, denoted as integer variable V, as shown in Constraint (18).

$$\sum_{i\in N_p} V_i = V \tag{18}$$

To specify the longest travel time of trips in group $r \in R$, we introduce binary variables A_t^r which are 1 when at least one trip in group $r \in R$ arrives at time $t \in T$, and 0 otherwise. Constraints (19) specify the arrival times of trips in group $r \in R$. Then, Constraints (20)–(22) calculate the longest travel time experienced by trips in group $r \in R$. Constraints (20) impose a lower bound to the longest travel time of trips in group $r \in R$ meaning that it has to be bigger than or equal to all of the different travel times experienced by travellers. Constraints (21) and (22) impose an upper bound to the longest travel time meaning that it has to be less than or equal to the longest travel time among all of the different travel times experienced by travellers in group $r \in R$. We define binary variables Z_t^r and impose that $\sum_{t|a^r+st^r\le t\le b^r} Z_t^r$ equals 1 to ensure that variable T^r_{AV} can only take one value which is the longest travel time of travellers using SAVs in group $r \in R$. $\mathcal{M}$ in Constraints (21) is a sufficiently large number.

$$\frac{E_t^r}{n^r} \le A_t^r \le E_t^r, \qquad \forall r \in R, a^r + st^r \le t \le b^r \tag{19}$$

$$T^r_{AV} \ge A_t^r (t - a^r), \qquad \forall r \in R, a^r + st^r \le t \le b^r \tag{20}$$

$$T_{AV}^{r} \le A_t^r(t-a^r) + \mathcal{M}(1-Z_t^r), \qquad \forall r \in R, a^r + st^r \le t \le b^r \tag{21}$$

$$\sum_{t|a^r+st^r \le t \le b^r} Z_t^r = 1, \qquad \forall r \in R \tag{22}$$

3.5. Traffic congestion

We include traffic congestion in the model by introducing flow-dependent travel time. It is important to note that only the flow of SAVs contributes to traffic congestion, as this study does not explore the mixed flow interaction between SAVs and bicycles. According to the BPR function (Dafermos and Sparrow, 1969), the travel time of traversing a road link has a non-linear relationship with the vehicle flow on this link: $t = t_0\left(1 + a\left(\frac{F}{Q}\right)^b\right)$. Here, a and b are parameters; t_0 represents the minimum travel time, F denotes the vehicle flow, and Q represents the link capacity. However, involving the non-linear BPR function makes the solving process more difficult. Therefore, we adopt the method proposed by Van Essen and Correia (2019) to replace the non-linear travel time calculation with multiple link-time-capacity choices. They use the concept of spatial link capacity $C_{i_{t_1}j_{t_2}}$ of a certain link $(i,j) \in L$ within a travel time slot between $t_1 \in T$ to $t_2 \in T$. The spatial link capacities can be calculated before the optimisation using the following equation.

$$C_{i_{t_1}j_{t_2}} = (t_2 - t_1)Q_{ij}\left(\frac{1}{a}\left(\frac{t_2-t_1}{t_{ij}^{\min}} - 1\right)\right)^{\frac{1}{b}} \tag{23}$$

We add 0.5 to t_2 when $t_2 - t_1$ equals $t_{ij}^{\min}$ to prevent the value of $C_{i_{t_1}j_{t_2}}$ from being zero. Among all the link-time-capacity choices, only one can be selected, meaning that there is a unique travel time for traversing road link (i,j) at a time instant $t_1 \in T$. This is described in Constraints (24) by making use of binary variables $X_{i_{t_1}j_{t_2}}$ which are 1 when any vehicle travels in road link (i,j) from time instant t_1 to t_2, where $(i_{t_1}, j_{t_2}) \in G$, and 0 otherwise. Note that using a large set of binary variables $X_{i_{t_1}j_{t_2}}$ in the time-space network may increase the complexity of solving the model (Kaufman et al., 1998). Specifying the set of binary variables requires the enumeration for each road link $(i,j) \in L$ at each time instant $t_1 \in T$ and $t_2 \in T$. However, in our case, given the maximum and minimum travel time of link $(i,j) \in L$, we only define binary variables for each link (i,j) from time instant $t_1 \in T$ to $t_2 \in T$ if $t_1 + t_{ij}^{\min} \le t_2 \le t_1 + t_{ij}^{\max}$. This reduces the number of binary variables required.

$$\sum_{t_1|\left(i_t, j_{t_1}\right) \in G} X_{i_t j_{t_1}} \le 1, \qquad \forall (i,j) \in L, t \in T \tag{24}$$

Constraints (25) require that the total flow on road link (i,j) from time instant t_1 to time instant t_2 never exceeds its corresponding spatial link capacity. Given that only one specific travel time will be chosen defined by Constraints (24), many flow variables $F_{i_{t_1}j_{t_2}}$ are imposed to zero.

$$F_{i_{t_1}j_{t_2}} \le \left\lfloor C_{i_{t_1}j_{t_2}} \right\rfloor X_{i_{t_1}j_{t_2}}, \qquad \forall \left(i_{t_1}, j_{t_2}\right) \in G \tag{25}$$

Vehicles that enter the road link first will leave the road link. This is known as the first-in-first-out (FIFO) rule, described by Constraints (26). These constraints only apply to time instant t_1 and t_2 when $t_1 < t_2 \le t_1 + t_{ij}^{\max} - t_{ij}^{\min}$. Otherwise, if $t_2 + t_{ij}^{\min} > t_1 + t_{ij}^{\max}$, there is no need to impose FIFO rule as vehicles entering the road link (i,j) at time instant t_1 with the longest travel time have left the link before any vehicles entering the road link (i,j) at a later time instant t_2 despite travelling with the shortest travel time.

$$t_1 + \sum_{t \in T} X_{i_{t_1}j_t}\left(t - t_1\right) \le t_2 + \sum_{t \in T} X_{i_{t_2}j_t}(t - t_2) + \mathcal{M}\left(1 - \sum_{t \in T} X_{i_{t_2}j_t}\right), \qquad \forall (i,j) \in L, \; t_1 < t_2 \le t_1 + t_{ij}^{\max} - t_{ij}^{\min} \tag{26}$$

3.6. Objective function

With the purpose of maximising the total profit of the SAV operator, we define the objective function as Eq. (27), which comprises the total revenue paid by the service users and the total costs of operating the whole system. Service users have to pay two types of fares for using an SAV, an initial fixed base fare p^0 in euros, and a distance-related price p in euros per kilometre. The distance-related price is charged based on the shortest travel distance sd^r of the trip $r \in R$ instead of the actual travel distance to avoid the unnecessary detours of SAVs in order to earn extra profits.

The total costs include the following: (1) the total depreciation costs of the SAV fleet in the system, which is calculated as the unit depreciation cost in euros per vehicle, denoted as cf, multiplied by the total fleet size. Here, cf is calculated as the vehicle's purchase price divided by its service life span; (2) the total operational costs including fuel, maintenance and insurance cost, which are calculated by multiplying the total distance of all the SAVs by the unit operational cost co in euros per kilometre; note that the total travel distance includes both the deliver distance with clients and the relocation distance without clients; (3) the penalty

for the late drop-off of the client, calculated by multiplying the delay cost cd in euros per time step by the difference between the actual riding time of clients and the shortest travel time.

$$\max \sum_{r\in R} OM^r_{AV} S^r - cf \cdot V - co\left(\sum_{\left(i_{t_1}, j_{t_2}\right)\in G} l_{ij} F_{i_{t_1} j_{t_2}}\right) - cd \sum_{r\in R}\left(\sum_{t\in T} tE^r_t - a^r S^r - st^r S^r\right) \tag{27}$$

where

$$OM^r_{AV} = p^0 + sd^r p, \qquad \forall r \in R. \tag{28}$$

4. Problem linearisation

The model proposed in Section 3 is a non-linear mixed integer programming model because of the exponential terms in binary logit model in Eqs. (4), the floor function to calculate the SAV demand in Eqs. (5) and the quadratic term to determine the acceptance rate in Constraint (6). To facilitate the solution process, we propose methods in Sections 4.1, 4.2, and 4.3 to linearise these non-linear equations and constraints, thereby transforming the model into a mixed integer linear programming model. In addition, we determine the most appropriate value for $\mathcal{M}$ used in Constraints (21) and (26) to get a tighter formulation of the proposed model, which is presented in Section 4.4.

4.1. Linearisation of the binary logit model

In Section 4.1.1, we first reformulate the binary logit model in Eqs. (4) with logarithmic functions which are still non-linear. Then, we adopt the outer-inner approximation method to linearise the logarithmic functions. Details of this method are described in Section 4.1.2. To use this method, a set of breakpoints needs to be pre-specified before the optimisation. Thus, in Section 4.1.3, we propose a breakpoint determination method to find the fewest breakpoints while guaranteeing a certain level of approximation accuracy.

4.1.1. Model reformulation

We firstly rewrite Eqs. (4) as:

$$\frac{P^r_{AV}}{1 - P^r_{AV}} = \frac{e^{V^r_{AV}}}{e^{V^r_B}}, \qquad \forall r \in R. \tag{29}$$

Then, we take the logarithm of both sides of Eqs. (29) to have the following equation:

$$\ln P^r_{AV} - \ln(1 - P^r_{AV}) = V^r_{AV} - V^r_B, \qquad \forall r \in R. \tag{30}$$

By defining new variables LN^r_{AV} and LN^r_B, we can further simplify the equation by having the following:

$$LN^r_{AV} = \ln P^r_{AV}, \qquad \forall r \in R, \tag{31}$$

$$LN^r_B = \ln(1 - P^r_{AV}), \qquad \forall r \in R, \tag{32}$$

$$LN^r_{AV} - LN^r_B = V^r_{AV} - V^r_B, \qquad \forall r \in R. \tag{33}$$

4.1.2. Linearisation of logarithmic functions: outer-inner approximation-based linear programming relaxation

We adopt the outer-inner approximation method (Wang et al., 2015; Guo et al., 2022) to linearise the logarithmic term in Constraints (31) and (32). The logarithmic function can be relaxed to a set of linear constraints that give the upper and the lower bound of the original logarithmic function (see Fig. 3). The procedure for linearising Constraints (31) and (32) are the same. For the sake of simplicity, we only take Constraints (31) as an example.

The original logarithmic function is divided into $\mathcal{K} - 1$ segments by $\mathcal{K}$ pre-determined breakpoints ($\mathcal{K}$ is the number of breakpoints). For each segment, the tangent lines at the two breakpoints and the secant line between the breakpoints serve as the upper bound and the lower bound of the real logarithm function, respectively. Note that the breakpoints can be distributed non-uniformly to minimise the approximation error. We introduce an index set for breakpoints, denoted by $K = \{1, 2, \dots, k, \dots, \mathcal{K}\}$. Each breakpoint $k \in K$ has coordinates $(u^k, \ln u^k)$ where $u^k \in [0, 1]$. The segment between two adjacent breakpoints $k \in K$ and $k + 1 \in K$ is denoted by $[u^k, u^{k+1}]$. If the value of variable P^r_{AV} falls within the interval $[u^k, u^{k+1}]$, we say that $[u^k, u^{k+1}]$ is an active interval. A binary variable λ^k_r is defined for $k \in \{1, 2, \dots, k, \dots, \mathcal{K}-1\}$, $r \in R$ to indicate whether or not an interval $[u^k, u^{k+1}]$ is active for group $r \in R$.

A set of constraints is introduced to describe this outer-inner approximation. Constraints (34) describe the tangent lines at each breakpoint, which serve as the outer approximation of the logarithmic function.

$$LN^r_{AV} \le \frac{1}{u^k} P^r_{AV} + \ln u^k - 1, \qquad \forall r \in R, k \in K \tag{34}$$

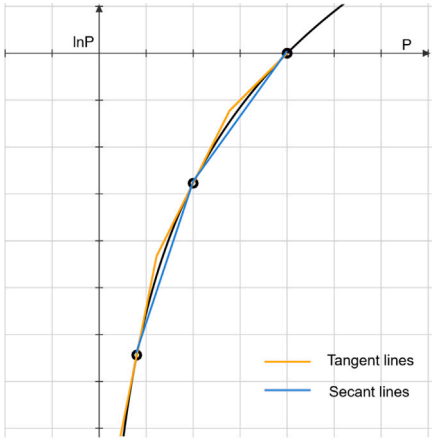


Fig. 3. Outer-inner approximation.

Constraints (35)–(41) describe the inner approximation of the logarithmic function. The value of variables LN^r_{AV} and P^r_{AV} can be represented by the convex combination of the coordinates of two consecutive breakpoints, where continuous variables θ^k_r are defined to represent the convex combination coefficient for breakpoint $k \in K$ for group of trips $r \in R$, as shown in Constraints (35) and (36).

$$LN^r_{AV} \geq \sum_{k=1}^{\mathcal{K}} \theta^k_r \ln u^k, \qquad \forall r \in R \tag{35}$$

$$P^r_{AV} = \sum_{k=1}^{\mathcal{K}} \theta^k_r u^k, \qquad \forall r \in R \tag{36}$$

The summation of coefficient θ^k_r has to be one, according to the convexity Constraints (37).

$$\sum_{k=1}^{\mathcal{K}} \theta^k_r = 1, \qquad \forall r \in R \tag{37}$$

There exists only one active interval, meaning that the value of P^r_{AV} and LN^r_{AV} can only fall into one line segment for each $r \in R$, ensured by Constraints (38).

$$\sum_{k=1}^{\mathcal{K}-1} \lambda^k_r = 1, \qquad \forall r \in R \tag{38}$$

The following constraints describe the relationship between two consecutive breakpoints and the active interval in between.

$$\theta^1_r \leq \lambda^1_r, \qquad \forall r \in R \tag{39}$$

$$\theta^k_r \leq \lambda^{k-1}_r + \lambda^k_r, \qquad \forall r \in R, k \in \{2, \ldots, \mathcal{K}-1\} \tag{40}$$

$$\theta^{\mathcal{K}}_r \leq \lambda^{\mathcal{K}-1}_r, \qquad \forall r \in R \tag{41}$$

4.1.3. Breakpoints determination

The approximation error from the outer-inner approximation can be reduced by using more breakpoints in the area where the nonlinear function has higher curvature. However, using too many breakpoints will significantly increase the number of variables and constraints, resulting in a heavy computational burden. In this section, we propose a breakpoint determination method with the aim of locating the fewest breakpoints with a good distribution so that a certain level of approximation accuracy can be guaranteed.

First of all, a maximum acceptable approximation (MAA) error for each interval needs to be specified as the threshold, denoted by γ. Then, the maximum approximation error between two consecutive breakpoints can be calculated given the equation of the logarithmic function and the approximation functions. Details on how to calculate the maximum approximation error are introduced later in this section. Next, specifying the coordinate of one breakpoint, the location of another breakpoint can be determined by ensuring that the maximum approximation error within the interval formed by these two breakpoints does not exceed the predetermined MAA error threshold γ. In this case, we can start from the last known breakpoint which is (1, 0) for the logarithmic

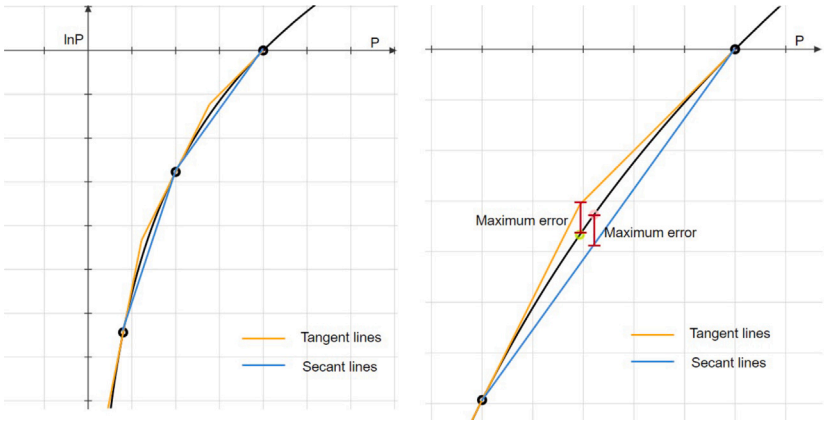


Fig. 4. Maximum approximation error.

function $\ln P_{AV}^r$, and then calculate the coordinate of the previous breakpoint. This procedure repeats until the x-coordinate of the newly found breakpoint is smaller than the lower bound of P_{AV}^r for all the groups $r \in R$, which is denoted as $\underline{P}$. $\underline{P}$ is defined as the minimum probability of choosing SAVs among all the groups, under which the demand for SAVs in all the groups will be zero. Namely, under $\underline{P}$, it is pointless to add additional breakpoints.

According to Constraints (5), when the value of $n^r P_{AV}^r$ is less than 0.5, the value of D_{AV}^r will be zero. Thus, $\underline{P}$ can take the largest value which satisfies Constraints (42).

$$\underline{P} \le \frac{1}{2n^r}, \qquad \forall r \in R \tag{42}$$

When the value of P_{AV}^r is less than $\underline{P}$, the demand for SAVs in group $r \in R$ will reach zero. However, this does not mean that P_{AV}^r cannot have a value less than $\underline{P}$. When the difference between the utility of SAVs and bicycles is sufficiently large, the probability of choosing SAVs may drop to near zero. To ensure the feasibility of the model, a boundary breakpoint must be added. The coordinates of this breakpoint can be specified as $(1/\mathcal{M}, \ln(1/\mathcal{M}))$, where $\mathcal{M}$ is a sufficiently large number.

As shown in Fig. 4, the approximated value lies between the tangent lines (yellow lines) and the secant lines (blue lines), while the real value is the logarithmic function (black line). Thus, the maximum approximation error is the maximum of (1) the maximum distance between the logarithmic function and the secant line, denoted as $e_1^{\max}$, and (2) the maximum distance between the logarithmic function and the tangent lines, denoted as $e_2^{\max}$. The maximum approximation error takes the maximum value between $e_1^{\max}$ and $e_2^{\max}$ which yields:

$$e^{\max} = \max\{e_1^{\max}, e_2^{\max}\}. \tag{43}$$

Proposition 4.1. *The maximum error between the logarithmic function and the secant line $e_1^{\max}$ equals the maximum error between the logarithmic function and the tangent lines $e_2^{\max}$. The maximum error at interval $[u^{k-1}, u^k]$ is:*

$$e^{\max} = \frac{u^{k-1}\ln u^k - u^k \ln u^{k-1}}{u^k - u^{k-1}} - \ln\left(\ln u^k - \ln u^{k-1}\right) + \ln\left(u^k - u^{k-1}\right) - 1. \tag{44}$$

Proof. We first calculate the maximum approximation error between the logarithmic function and the secant line.

For $P_{AV}^r \in [u^{k-1}, u^k]$, we define the error at P_{AV}^r between the approximated value and the real logarithmic function as e, where

$$e = \ln P_{AV}^r - \left(\ln u^{k-1} + \frac{\ln u^k - \ln u^{k-1}}{u^k - u^{k-1}}\left(P_{AV}^r - u^{k-1}\right)\right). \tag{45}$$

To determine which point between these two breakpoints contributes to the maximum error, we have to set the derivative of the error e to 0:

$$\frac{1}{P_{AV}^r} - \frac{\ln u^k - \ln u^{k-1}}{u^k - u^{k-1}} = 0. \tag{46}$$

This results in

$$P_{AV}^r = \frac{u^k - u^{k-1}}{\ln u^k - \ln u^{k-1}}. \tag{47}$$

At this point, the maximum approximation error occurs. The approximation error equals the difference between the logarithmic function and the secant line at this point.

$$e_1^{\max} = \ln\left(u^k - u^{k-1}\right) - \ln\left(\ln u^k - \ln u^{k-1}\right) + \frac{u^{k-1}\ln u^k - u^k \ln u^{k-1}}{u^k - u^{k-1}} - 1. \tag{48}$$

The maximum approximation error between the tangent lines and the logarithmic function occurs at the point where two tangent lines of the consecutive breakpoints intersect. Knowing the coordinates of the two consecutive breakpoints $(u^k, \ln u^k)$ and $(u^{k-1}, \ln u^k)$ with $u^{k-1} < u^k$, the tangent lines at those two breakpoints can be expressed as follows:

$$y^k = \frac{1}{u^k}x + \ln u^k - 1, \tag{49}$$

$$y^{k-1} = \frac{1}{u^{k-1}}x + \ln u^{k-1} - 1. \tag{50}$$

Combining the two equations, we can calculate the intersection point of the two tangent lines, which is

$$\left(\frac{(\ln u^k - \ln u^{k-1})u^k u^{k-1}}{u^k - u^{k-1}}, \frac{(\ln u^k - \ln u^{k-1})u^{k-1}}{u^k - u^{k-1}} + \ln u^k - 1\right).$$

The maximum approximation error $e_2^{\max}$ is the vertical distance from the intersection point to the logarithmic function:

$$e_2^{\max} = \frac{\left(\ln u^k - \ln u^{k-1}\right)u^{k-1}}{u^k - u^{k-1}} + \ln u^k - 1 - \ln\left(\frac{\left(\ln u^k - \ln u^{k-1}\right)u^k u^{k-1}}{u^k - u^{k-1}}\right). \tag{51}$$

This gives:

$$e_2^{\max} = \frac{u^{k-1}\ln u^k - u^k \ln u^{k-1}}{u^k - u^{k-1}} - \ln\left(\ln u^k - \ln u^{k-1}\right) + \ln\left(u^k - u^{k-1}\right) - 1. \tag{52}$$

So, we have $e^{\max} = e_1^{\max} = e_2^{\max}$, with the maximum approximation errors occurring at different locations. □

Setting the last breakpoint equal to $(1, 0)$ and fixing the desired maximum error γ, we can determine the breakpoint before 1 by solving the following formula numerically for u^{k-1}.

$$\gamma = \ln\left(u^k - u^{k-1}\right) - \ln\left(\ln u^k - \ln u^{k-1}\right) + \frac{u^{k-1}\ln u^k - u^k \ln u^{k-1}}{u^k - u^{k-1}} - 1 \tag{53}$$

with

$$u^k = 1. \tag{54}$$

Similarly, u^{k-2} can be obtained by using the found u^{k-1} as input.

4.2. Linearisation of the floor function

The demand calculation function in Eqs. (5) is non-linear. Therefore, we replace Eqs. (5) by the following constraints:

$$n^r \cdot P_{AV}^r - 0.5 < D_{AV}^r \leq n^r P_{AV}^r + 0.5, \qquad \forall r \in R. \tag{55}$$

4.3. Linearisation of the acceptance rate constraint

When an SAV operator is allowed to reject nonprofitable requests, α is defined as a continuous variable with $\alpha \in [0, 1]$. As a result, Constraint (6) becomes a non-linear constraint consisting of the product of the continuous variable α and the integer variables D_{AV}^r. To linearise this constraint, we introduce additional binary variables $\overline{D}_h$ to discretise the integer term $\sum_{r\in R} D_{AV}^r$, and continuous variables $Y_h \in [0, 1]$ to describe the value of the integer term $\sum_{r\in R} S^r$, where $h \in \{0, 1, \ldots, \mathcal{H}\}$. Here, $\mathcal{H}$ should be chosen such that these constraints still hold when all demand would use SAVs.

Then, we substitute Constraint (6) with the following constraints.

$$\sum_{r\in R} D_{AV}^r = \sum_{h=0}^{\mathcal{H}} 2^h \overline{D}_h \tag{56}$$

$$Y_h \leq \alpha, \qquad \forall h \in \{0, 1, \ldots, \mathcal{H}\} \tag{57}$$

$$Y_h \leq \overline{D}_h, \qquad \forall h \in \{0, 1, \ldots, \mathcal{H}\} \tag{58}$$

$$Y_h \geq \alpha + \overline{D}_h - 1, \qquad \forall h \in \{0, 1, \ldots, \mathcal{H}\} \tag{59}$$

$$\sum_{h=0}^{\mathcal{H}} 2^h Y_h = \sum_{r\in R} S^r \tag{60}$$

4.4. Tightening the model by choosing an appropriate value for $\mathcal{M}$

$\mathcal{M}$ used in Constraints (21) and (26) represents a sufficiently large number. However, using an excessively large $\mathcal{M}$ may lead to a model with a weak relaxation, which can, in turn, slow down the solving process of the mixed-integer programming (MIP) model. Thus, choosing a proper value for $\mathcal{M}$ is beneficial in tightening the proposed model. The main criterion for choosing an appropriate value for $\mathcal{M}$ is to identify the smallest value that is sufficiently large to prevent the cut-off of any feasible solution. The value of $\mathcal{M}$ should be specified for each of the constraints to get a tighter formulation.

We first rewrite Constraints (21) as follows with constraint-specific values $\mathcal{M}_r^1$ where $r \in R$.

$$T_{AV}^r \le A_t^r(t - a^r) + \mathcal{M}_r^1(1 - Z_t^r), \qquad \forall r \in R, a^r + st^r \le t \le b^r \tag{61}$$

Given that the longest travel time T_{AV}^r for SAV in group $r \in R$ is inherently less than or equal to the time difference between the latest arrival time b^r and the departure time a^r, $\mathcal{M}_r^1$ can take the following value regardless of the values of the binary variables A_t^r.

$$\mathcal{M}_r^1 = b^r - a^r, \qquad \forall r \in R \tag{62}$$

Then, we rewrite Constraints (26) using a constraint-specific value $\mathcal{M}_{ijt_1t_2}^2$ with $t_1, t_2 \in T$ if $t_1 < t_2 \le t_1 + t_{ij}^{\max} - t_{ij}^{\min}, (i,j) \in L$.

$$t_1 + \sum_{t \in T} X_{i_{t_1} j_t}\left(t - t_1\right) \le t_2 + \sum_{t \in T} X_{i_{t_2} j_t}(t - t_2) + \mathcal{M}_{ijt_1t_2}^2\left(1 - \sum_{t \in T} X_{i_{t_2} j_t}\right), \qquad \forall (i,j) \in L, \; t_1 < t_2 \le t_1 + t_{ij}^{\max} - t_{ij}^{\min} \tag{63}$$

When the value of $\sum_{t \in T} X_{i_{t_2} j_t}$ is 0, indicating that no vehicles enter the link $(i,j) \in L$ at time instant $t_2 \in T$, Constraints (63) become the follows:

$$t_1 + \sum_{t \in T} X_{i_{t_1} j_t}\left(t - t_1\right) \le t_2 + \mathcal{M}_{ijt_1t_2}^2, \qquad \forall (i,j) \in L, t_1 < t_2 \le t_1 + t_{ij}^{\max} - t_{ij}^{\min}. \tag{64}$$

The left-hand side of Constraints (64) indicates the time that a vehicle leaves link $(i,j) \in L$ if it enters this link at time instant $t_1 \in T$. Knowing that a maximum travel time for a vehicle traversing link $(i,j) \in L$ is $t_{ij}^{\max}$, the latest time that a vehicle leaves link $(i,j) \in L$ can never exceed its maximum travel time plus the entering time, which gives:

$$t_1 + \sum_{t \in T} X_{i_{t_1} j_t}\left(t - t_1\right) \le t_1 + t_{ij}^{\max}, \qquad \forall t_1 \in T, (i,j) \in L. \tag{65}$$

Combining Constraints (64) and (65) gives the smallest value that $\mathcal{M}_{ijt_1t_2}^2$ can take, which is:

$$\mathcal{M}_{ijt_1t_2}^2 = t_1 + t_{ij}^{\max} - t_2, \qquad \forall (i,j) \in L, t_1 < t_2 \le t_1 + t_{ij}^{\max} - t_{ij}^{\min}. \tag{66}$$

To help readers comprehend the model more efficiently, we summarise the complete problem formulation as well as the notations of the sets, parameters, and variables in Appendix.

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

In this section, we present the computational results of the case study of Delft, in The Netherlands to evaluate the effectiveness of the proposed model.

5.1. Application setting

The proposed model is applied to a quasi-real case study of the city of Delft, in the South Holland province in The Netherlands (Correia and Van Arem, 2016). A simplified road network of Delft is used in this case study which contains 35 nodes and 104 directed links (two-way circulation allowed), displayed in Fig. 5. SAVs are free to drive on the entire network, but only 7 nodes are designated as free parking depots, which are nodes 3, 10, 11, 15, 19, 22, and 27 (identified in red). The parking depots are distributed throughout the city, with three located in the city centre and four located on the outskirts, which facilitates the use of SAV services by residents from all areas of the city. In addition, each road link has either one lane or two lanes, with a capacity of 1600 or 3200, respectively. Vehicles are allowed to travel on these two types of road links with a maximum travel speed (free-flow speed without congestion) of 50 km/h and 70 km/h, respectively, and with a minimum travel speed of 5 km/h.

The mobility data for the morning peak hour in Delft was obtained using the Dutch mobility dataset (MON 2007/2008). This dataset provides the daily mobility information of a sample of residents, including but not limited to the origin, destination, departure time, arrival time, transport mode, etc. It has been used previously (Correia and Van Arem, 2016; Liang et al., 2020) to study the future mobility system with AVs in urban networks. However, this dataset does not have a large sample of trips for this city if we focus on just one hour. To overcome this limitation and characterise as much as possible the real mobility pattern in the morning peak hour, we filtered out the trips in the database from 7 am to 10 am with travel modes of bicycle, car, and taxi, and then evenly distributed them within one hour with one-third of the amount. In total, 2933 trips are generated, aggregated into 45 groups of trips by the similarity of the trip information as explained in the model.

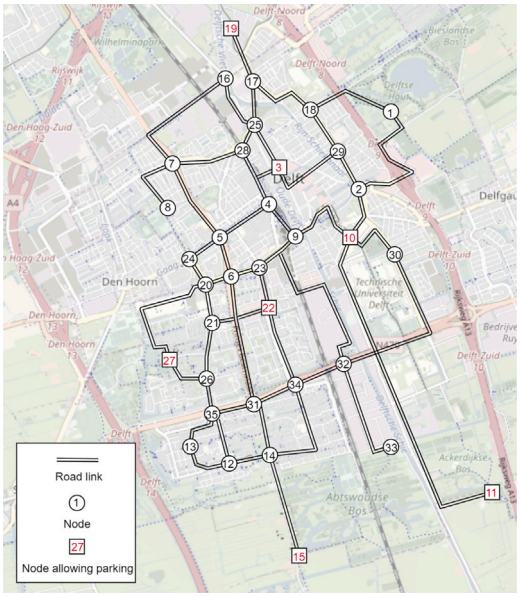


Fig. 5. Simplified road network of Delft used in the case study.

Table 2
Parameter summary.

Parameter	Description and reference values
$t_{ij}^{\max}$	Maximum travel time by SAVs which is computed by dividing the length of the road link by the minimum travel speed of 5 km/h.
$t_{ij}^{\min}$	Minimum travel time by SAVs which is calculated by dividing the length of the road link by the related maximum travel speed (50 km/h or 70 km/h). Note that the minimum travel time on each road link has a minimum value of 1 time step (2.5 min in this case study) due to the time-space network. It imposes that no vehicles can travel with a travel time of zero.
$C_{i_{t_1} j_{t_2}}$	Spatial capacity of each road link which is calculated using Eq. (23).
sd^r, st^r	Shortest travel distance/time which is calculated using the shortest path algorithm assuming SAVs can travel with free-flow speed.
T_B^r	Travel time of bicycles which is calculated by dividing the length of the shortest path by the average speed of the Dutch on a pedal bicycle, 12.4 km/h (BicycleDutch, 2018).
β_0	Parameter used in the logit model with a value of 0.1.
VOT_{AV}^r	Travellers' VOTT for using an AV with high income, middle income, and low income equal to 6.6, 4.6, and 3.8 euro/h, respectively (Kolarova et al., 2019).
VOT_B^r	Travellers' VOTT for using a bicycle with high income, middle income, and low income equal to 24.9, 17.3, and 14.1 euro/h, respectively (Kolarova et al., 2019).
p^0, p	Initial base fare and price per km which are set to 3 euros and 1.28 euros/km, respectively, according to the price rate of Uber in Delft, in The Netherlands (Uber, 2023).
co	Operational cost of SAVs which is set to 0.32 euro/km (calculated according to the methodology proposed by Bösch et al., 2018).
cf	Depreciation cost of SAVs which is set to 1.2 euro/vehicle/h (Fan et al., 2022).
cd	Delay penalty which is set to 0.2 euro/min (Liang et al., 2020).
α	Service rate which is set to 1 when the SAV operator has to serve all the trips.
a, b	Parameters in the BPR function which are set to 2 and 4, respectively (Van Essen and Correia, 2019).

The optimisation period contains two parts. One is a one-hour period studied during the morning peak, comprised of 24 time steps of each 2.5 min. Besides that, 5 additional time steps are added as a pre-optimisation period. This is needed since we assume that all the SAVs depart from parking depots in the morning to serve the trips. For SAVs to arrive at the requested origin on time, additional time steps are required as slack in the optimisation period. Therefore, the optimisation period contains a total of 29 time steps.

The parameters used in this case study related to the network setting, the demand, and the SAV operating system are summarised and explained in Table 2.

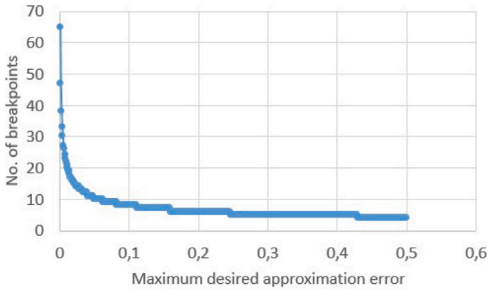


Fig. 6. Relationship between MAA error and the number of generated breakpoints.

Table 3
Optimisation results with different MAA errors.

MAA error	0.005	0.01		0.05	
	Value	Value	Relative change	Value	Relative change
Number of breakpoints	31	23	−28.57%	11	−64.29%
Objective function value	11 128.07	11 143.30	+0.14%	11 414.46	+2.57%
Optimal fleet size of SAVs	890	891	+0.11%	914	+2.70%
Demand for SAVs	1269	1271	+0.16%	1304	+2.76%
Computational time	3126 s	2236 s	−28.47%	1921 s	−38.55%

5.2. Breakpoint generation

Before solving the reformulated MILP model, a set of breakpoints was generated with a pre-specified MAA error. This error represents the maximum acceptable difference between the approximation value and the true value of the non-linear terms in Constraints (31) and (32). Therefore, the smaller this value is, the more precise the approximation will be; however, the greater the number of breakpoints that will be generated.

Fig. 6 shows the relationship between the number of generated breakpoints and the value of the MAA error. We can observe a clear trend where the number of generated breakpoints increases dramatically with the decrease in the value of the MAA error, especially when the approximation error is less than 0.05. On the one hand, more breakpoints lead to higher accuracy, but on the other hand, they lead to greater computational time. To balance these two factors, we need an MAA error that can ensure a good quality of the optimisation results within an acceptable computational time. To find a proper value for this case study, we first tested the model in the base scenario with three different values for the MAA error, which are 0.05, 0.01, and 0.005, yielding 11, 23, and 31 breakpoints, respectively.

We implemented the reformulated MILP model in Python 3.7 and then solved it using Gurobi optimiser version 10.0.0 on an Intel(R) Xeon(R) W-2123 CPU @3.60 GHz, and 32.00 GB RAM computer. Table 3 shows the optimisation results with the three mentioned MAA errors. Here, we used the objective function value, the optimal total fleet size, the total demand for SAVs, and the computational time as indicators to compare the performance for these three cases. The objective function value is the maximised profit for an SAV operator. The optimal fleet size and the total demand for SAVs are selected as the indicators because fleet sizing is one of the most important planning decisions for an SAV operator, and the demand for SAVs directly impacts the fleet sizing decision. In addition to this reason, the demand for SAVs is one of the attributes most affected by the approximation error.

We first ran the model with MAA error of 0.005, then used the corresponding optimisation results as the benchmark to compare with other cases in which the MAA error is 0.01 and 0.05. The relative changes in the values of the indicators are computed. Looking at the optimisation results with MAA values of 0.005 and 0.01 in Table 3, we observe very small differences in the objective function values (0.14% relative difference), the values of the optimal fleet sizes (1 unit difference), and the values of the demand for SAVs (2 units difference), indicating that using 23 breakpoints has already achieved a good approximation accuracy. Adding more breakpoints does not bring a significant improvement to the optimisation outcomes.

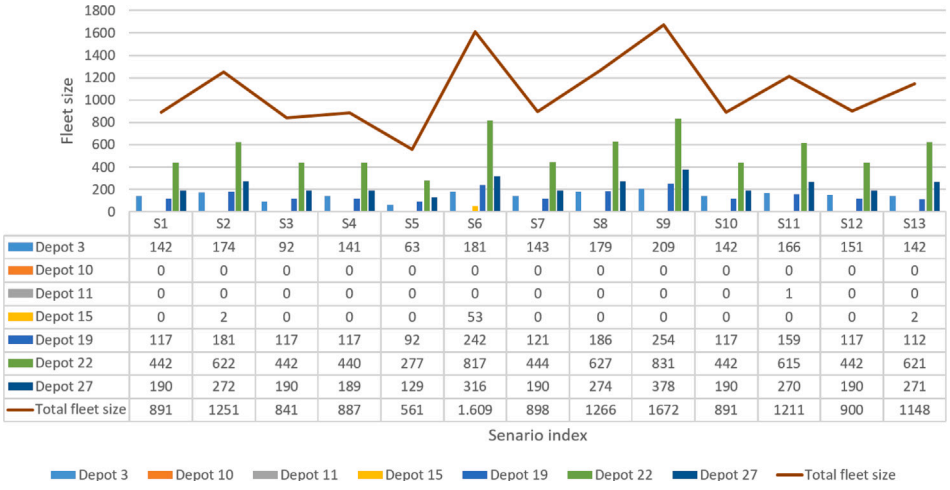
In all, the MAA error of 0.01 was used throughout the experiments, which yields 23 breakpoints.

5.3. Optimisation results

The model was tested first in a base scenario with parameters given in Section 5.1. Then, we conducted a sensitivity analysis to the following parameters: SAVs price rate, unit operational cost, delay penalty, parameter β_0, and the combination of them in 6 scenarios. We also investigate the impact of congestion by evaluating non-congested scenarios as a comparison of the existing scenarios. As previously mentioned, this paper explores two accept/reject mechanisms. Scenarios 1 to 9 assume that the SAV operator must accept all requests, while scenarios 10 to 13 assume that the SAV operator may reject the nonprofitable trips. A sensitivity analysis of SAV price rate and parameter β_1 is carried out to see how travellers' satisfaction with the service quality level influences the managerial decisions of the SAV operator under different pricing policies.

Table 4
Scenario description.

Scenario description	p^0 (euro)	p (euro/km)	co (euro/km)	α	cd (euro/min)	β_0 (utility/euro)	β_1
S1 Base scenario	3	1.28	0.32	1	0.2	0.1	–
S2 Lower price	1.5	0.64	0.32	1	0.2	0.1	–
S3 Lower operational cost	3	1.28	0.1	1	0.2	0.1	–
S4 No delay penalty	3	1.28	0.32	1	0	0.1	–
S5 Higher β_0	3	1.28	0.32	1	0.2	0.5	–
S6 Higher β_0 with lower price	1.5	0.64	0.32	1	0.2	0.5	–
S7 Base scenario without congestion	3	1.28	0.32	1	0.2	0.1	–
S8 Lower price without congestion	1.5	0.64	0.32	1	0.2	0.1	–
S9 Higher β_0 with lower price without congestion	1.5	0.64	0.32	1	0.2	0.5	–
S10 Base scenario with rejection	3	1.28	0.32	–	0.2	0.5	1
S11 Lower price with rejection	1.5	0.64	0.32	–	0.2	0.5	1
S12 Base scenario with rejection and lower β_1	3	1.28	0.32	–	0.2	0.5	0.1
S13 Lower price with rejection and lower β_1	1.5	0.64	0.32	–	0.2	0.5	0.1



	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
Depot 3	142	174	92	141	63	181	143	179	209	142	166	151	142
Depot 10	0	0	0	0	0	0	0	0	0	0	0	0	0
Depot 11	0	0	0	0	0	0	0	0	0	0	1	0	0
Depot 15	0	2	0	0	0	53	0	0	0	0	0	0	2
Depot 19	117	181	117	117	92	242	121	186	254	117	159	117	112
Depot 22	442	622	442	440	277	817	444	627	831	442	615	442	621
Depot 27	190	272	190	189	129	316	190	274	378	190	270	190	271
Total fleet size	891	1251	841	887	561	1.609	898	1266	1672	891	1211	900	1148

Fig. 7. Fleet size and initial distribution of SAVs at the beginning of a day in all the scenarios.

Table 5
Optimisation results for all the scenarios.

Scenario	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
Total profit (euro)	11 143.31	3389.07	13 904.24	11 695.33	7530.06	3653.06	11 924.48	4905.77	6437.42	11 143.31	3463.92	11 152.51	3505.19
Total revenue (euro)	16 663.61	11 729.52	16 672.16	16 601.24	10 846.61	15 113.09	16 836.59	11 874.50	15 667.01	16 663.61	11 159.4	16 599.62	10 691.84
Average price per trip (euro)	13.11	6.84	13.11	13.10	14.23	7.07	13.12	6.84	7.02	13.11	6.92	13.15	6.81
Total depreciation cost (euro)	1069.20	1501.20	1009.20	1064.40	673.20	1930.80	1077.60	1519.20	2006.40	1069.20	1453.2	1080.0	1377.6
Total operational cost (euro)	3962.10	5765.25	1285.72	3841.51	2559.35	7510.24	3834.51	5449.54	7223.18	3962.10	5485.27	3934.11	5245.55
Total delay penalty cost (euro)	489	1074	473	0	84	2019	0	0	0	489	757.0	433	563.5
Total demand for SAVs	1271	1715	1272	1267	762	2138	1283	1737	2232	1271	1690	1273	1714.0
SAV demand share	43.33%	58.47%	43.37%	43.20%	25.98%	72.89%	43.74%	59.22%	76.10%	43.33%	57.62%	43.4%	58.44%
Total satisfied trips for SAVs	1271	1715	1272	1267	762	2138	1283	1737	2232	1271	1613	1262	1570
Percentage of satisfied demand	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	95.44%	99.14%	91.6%
Average delay per trip (min)	1.93	3.13	1.85	3.25	0.55	4.73	0	0	0	1.93	2.35	1.68	1.8
SAVs total travel distance (km)	12 381.57	18 016.39	12 857.21	12 004.73	7997.96	23 469.50	11 982.85	17 029.80	22 572.45	12 381.57	17 141.48	12 294.10	16 392.34
SAVs total deliver distance (km)	10 552.54	15 554.53	10 807.14	10 324.48	6843.03	20 266.68	10 146.56	14 482.82	19 248.45	10 552.54	14 907.98	10 523.65	14 041.77
SAVs total relocate distance (km)	1829.03	2461.86	2050.06	1680.25	1154.92	3202.82	1836.29	2546.98	3324	1829.03	2233.49	1770.46	2350.57
SAVs total delivery time (h)	339.5	512.08	338.29	366.25	201.75	715.04	301.5	427.33	567.21	339.5	467.96	334.08	435.54
Average delivery time per trip (min)	16.03	17.93	15.95	17.35	15.88	20.08	14.1	14.75	15.25	16.03	17.4	15.88	16.65
Computational time (s)	1986 s	7846 s	2398 s	91 545 s	634 s	86 400 s	337 s	386 s	388 s	9517 s	60 758 s	6145 s	11 510 s
MIP Gap	0	0	0	0	0	0.57%	0	0	0	0	0	0	0

Table 4 shows the descriptions and parameter settings for all scenarios. The optimal fleet size distribution can be found in Fig. 7 and key performance indicators can be found in Table 5.

5.3.1. Base scenario

As can be seen in Fig. 7, almost all the SAVs are distributed at parking depots 3, 19, 22, and 27 at the beginning of the day, as these depots are either close to residential areas or the train station in Delft where the commuting demand is high during the morning peak hour. Depots 10, 11, and 15 have hardly any SAVs as these depots are either located on the outskirts of the city or near the campus area which is usually the destination of commuting in the morning. The distribution of the fleet at the beginning of the day is highly influenced by the geographical distribution of the population, the distribution of land use, and the travel patterns of residents.

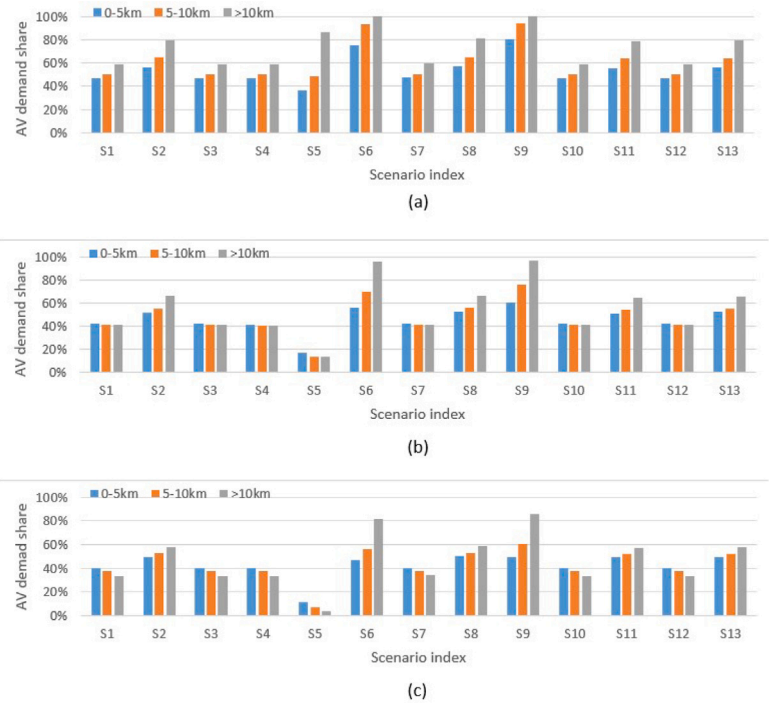


Fig. 8. SAV demand share for different user classes with (a) high VOTT, (b) middle VOTT, and (c) low VOTT.

Table 6
Optimisation results under the base scenario.

User class		0–5 km	5–10 km	≥10 km
High VOTT	SAV demand share	46.90%	50.23%	59.22%
	Average price per trip	8.04	12.32	20.81
	Average travel time-related cost per trip for using an SAV	0.99	1.79	2.70
	Average travel time-related cost per trip for using a bicycle	7.53	13.82	27.03
Middle VOTT	SAV demand share	41.59%	40.93%	40.78%
	Average price per trip	8.04	12.32	20.81
	Average travel time-related cost per trip for using an SAV	0.76	1.36	1.98
	Average travel time-related cost per trip for using a bicycle	5.23	9.63	18.78
Low VOTT	SAV demand share	40.00%	37.33%	33.33%
	Average price per trip	8.04	12.32	20.81
	Average travel time-related cost per trip for using an SAV	0.62	1.07	1.75
	Average travel time-related cost per trip for using a bicycle	4.27	7.85	15.31

In Table 5, 43.33% of the travellers choose to use SAV services. However, travellers from different income classes behave differently facing trips with different lengths, as shown in Table 6. We classified trips into three groups in terms of their lengths: less than 5 km, between 5 and 10 km, and more than 10 km. Then, we calculated the demand share of SAV services for travellers in different classes (with high VOTT, middle VOTT, and low VOTT), and their corresponding cost structures (price, travel time-related cost for using an SAV, and a bicycle).

Results indicate that travellers with a high VOTT are more sensitive to variations in trip length compared with the other classes. When the trip length is longer than 10 km, 59.22% of travellers with a high VOTT use SAVs rather than cycling because the increase in time-related costs of cycling is significant for them. When the length of the trip is short (less than 5 km), 46.9% of travellers with a high VOTT choose SAVs, meaning that using bicycles can slightly save them some costs. But the difference between using these two modes is not big. For trips between 5 km and 10 km, half of the people choose SAVs as the utilities for using these two modes are the same. It makes little difference which mode they choose. Note that travellers with a middle VOTT and a low VOTT always prefer bicycles to SAVs as the price for using SAVs is high. In addition, travellers with a middle VOTT are insensitive to the changes in trip length. For them, the cost difference between these two modes does not change significantly with the increase in trip length.

5.3.2. Sensitivity analysis on price rate

The price rate has a great impact on travellers' behaviour, which in turn affects the total demand for SAV services and fleet sizing decisions. In Table 5, one can see that the demand for SAVs increased from 1271 (in S1 Base scenario) to 1715 (in S2 Lower price)

when the price rate reduces to half of what it was in the base scenario S1. To satisfy the increased demand, the SAV operator has to deploy a larger fleet size. At the same time, the congestion level increased as more vehicles circulated on the road network and competed for the shortest paths. It can be seen from Table 5 that the average delay time per trip increased from 1.93 (in S1 Base scenario) to 3.13 min (in S2 Lower price) and the average delivery time per trip increased from 16.03 min (in S1 Base scenario) to 17.93 min (in S2 Lower price). Thus, more delay penalty was generated which reduced the total profits. Even though more travellers choose to use SAVs, the total profit is still lower than the one in S1 Base scenario because of the lower revenue, higher depreciation costs, higher operational costs and higher delay penalty.

When the price rate for using an SAV is reduced, the willingness of travellers from different income classes to use the SAV service increases compared with S1 Base scenario, as can be seen in Fig. 8. For travellers with high VOTT (in Fig. 8a), SAVs are preferable to bicycles regardless of trip length. For travellers with middle VOTT whose trips are less than 10 km in length, the preference between the two modes is not obvious. Only when the trip length exceeds 10 km, do travellers prefer to use SAVs over the bicycle mode. A similar trend is observed for travellers with a lower VOTT. The longer the trip is, the more likely a traveller with a low VOTT will choose SAVs.

5.3.3. Sensitivity analysis on operational costs

As depicted in Fig. 7, the total fleet size in S3 Lower operational cost is 50 vehicles smaller than the one in S1 Base scenario. Looking at the optimisation results displayed in Table 5, we found that the total profit increases from 11 143.31 euros (in S1 Base scenario) to 13 904.24 euros (in S3 Lower operational cost). The change in demand for SAVs in S1 and S3 is negligible, and the total relocation distance increases from 1829.03 km (in S1 Base scenario) to 2050.06 km (in S3 Lower operational cost). This indicates that the SAV operator can save money by deploying a smaller fleet and allowing SAVs to relocate more frequently at a lower operational cost.

In addition, the total delay penalty cost decreases from 489 euros (in S1 Base scenario) to 473 euros (in S3 Lower operational cost), while the total delivery distance of the SAVs increases from 10 552.54 km (in S1 Base scenario) to 10 807.14 km (in S3 Lower operational cost), meaning that SAVs detour more to avoid traffic congestion to deliver clients as soon as possible, as well as to pick up more clients.

Looking at Fig. 8, we barely notice any difference between the SAV shares for different user classes and trip length. Thus, we conclude that lowering the operational cost of the SAV fleet does not have a significant influence on the demand structure.

All in all, SAV operators earn more profits through operational cost savings, less delay penalty, and fewer depreciation costs of the fleet, even though SAVs detour and relocate more.

5.3.4. Sensitivity analysis on delay penalty

The SAV operator earns greater profits when there is no delay penalty for the late drop-off of clients in S4. However, the attractiveness of the SAV service drops slightly, which is reflected in the reduced demand for SAVs, which can be seen in Table 5. Furthermore, the reduced amount of fleet size is consistent with the decreased demand for SAVs.

To compare the trip delay information between S4 (no delay penalty) and S1 (base scenario), we have plotted the delay distributions, along with the mean and the 90th percentile values for the delay in Fig. 9. The results indicate that 90% of trips in S1 experience a delay within 5 min, whereas 90% of trips in S4 have a delay within 7.5 min. Additionally, the average delay per trip is 1.93 min in S1 and 3.25 min in S4. These findings indicate that when there is no penalty for late deliveries, the actual delivery time becomes longer compared to the base scenario S1, resulting in increased delay, which can also be seen from the increased average delivery time per trip and the increased average delay per trip in S4 in Table 5. However, it is crucial to consider the perspective of passengers using the SAV service in an inter-modal fashion, who require a certain level of reliability in their arrival time, because of the need to coordinate with other transportation modes. In scenarios where an SAV cannot guarantee arrival before a traveller's acceptable latest arrival time, the trip may be rejected, and the traveller may opt for an alternative mode of transportation. From this perspective, implementing a delay penalty for the SAVs could encourage more reliable and timely deliveries, addressing passengers' concerns and potentially reducing trip rejections. The value of the delay penalty will influence the demand for SAVs, and consequently, impact overall profitability. Investigating the optimal delay penalty value remains an area for future research.

The value of the delay penalty will influence the demand for Shared Autonomous Vehicles (SAVs), and consequently, impact overall profitability. Investigating the optimal delay penalty value and the spatial impact of the rejection rate remains an area for future research.

Overall, the delay penalty does not have a significant impact on fleet sizing decisions and travellers' behaviour. Travellers with higher VOTT care more for late arrival than those with a relatively lower VOTT.

5.3.5. Sensitivity analysis on β_0

The parameter β_0 indicates the level of sensitivity that travellers exhibit towards the changes in monetary costs. In this section, we tested a higher value of β_0 which is 0.5 utility/euro in two scenarios S5 and S6, based on S1 Base scenario and S2 Lower price. We tested these two scenarios with different price rates because the congestion level is different in both, which allows for observing the impact of network congestion levels on the optimisation results.

Looking at Fig. 7, we found a big difference in the optimal fleet size in S5 Higher β_0 and S6 Higher β_0 with lower price, compared with S1 Base scenario and S2 Lower price. The fleet size differences mainly come from the decreased/increased demand in these two scenarios, which are 762 and 2138, as can be found in Table 5. When looking into the details of SAV demand share in S5

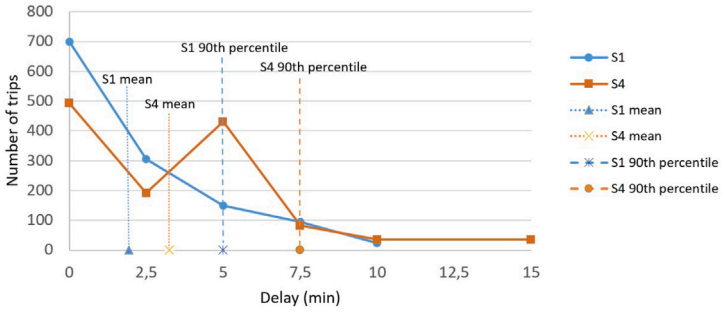


Fig. 9. Delay distribution in S1 (base scenario) and S4 (no delay penalty).

Higher β_0 and S6 Higher β_0 with lower price, we notice that more travellers tend to choose the mode with the least generalised costs, resulting in a greater difference between the demand share for SAVs and bicycles. Note that the variation of the SAV demand share for all the user classes in all types of trips in S5 and S6 share a similar trend as in S1 and S2, indicating that a higher value of β_0 can only bring a larger degree of variation to the demand share, but it cannot completely alter travellers' preference towards the modes.

A value of β_0 that is precisely estimated for the application city can enhance the realism of the model and lead to a more accurate fleet sizing decision. Note that the model was solved in S6 Higher β_0 with a lower price with a MIP gap of 0.57% with a time limit of 24 h because the congestion level is high due to the increased demand for SAVs.

5.3.6. *Impact of traffic congestion*

To test the impact of traffic congestion on the strategic and operational decisions, we removed traffic congestion from S1 Base scenario, S2 Lower price and S6 Higher β_0 with lower price, by assuming all vehicles can travel at free-flow speed. These three scenarios are selected as references because they exhibit gradually increased congestion levels. Removing congestion in these three scenarios gives us three new scenarios named S7, S8, and S9.

In Fig. 7, it can be seen that the optimal fleet size increases from 891 (in S1 Base scenario) to 898 (in S7 Base scenario without congestion), from 1251 (in S2 Lower price) to 1266 (in S8 Lower price without congestion), from 1609 (in S6 Higher β_0 with lower price) to 1670 (in S9 Higher β_0 with a lower price without congestion). It turns out that congestion has a significant impact on fleet sizing decisions, which should be taken into consideration when solving the fleet management problem.

Without congestion, all trips can be delivered to the desired destinations in the shortest travel time and travel distance, as can be seen in Table 5. The average delay per trip and the total delay penalty in S7 Base scenario without congestion, S8 Lower price without congestion and S9 Higher β_0 with lower price without congestion are 0. Consequently, the demand for SAVs increased from 1271 (in S1 Base scenario) to 1283 (in S7 Base scenario without congestion), from 1715 (in S2 Lower price) to 1737 (in S8 Lower price without congestion), and from 2138 (in S6 Higher β_0 with lower price) to 2232 (in S9 Higher β_0 with lower price without congestion) because travellers are more willing to take SAVs if the travel time is lower. However, despite an increase in demand, the total travel distance and the total delivery time of SAVs decrease correspondingly, indicating that SAVs no longer need to take longer detours to avoid the competition for the shortest paths, which reduces operational costs significantly.

5.3.7. *Comparison of the two accept/reject mechanisms*

In terms of the accept/reject mechanism, from S10 to S13, the SAV operator can reject nonprofitable trips. α is defined as a continuous variable that represents the trip service rate. However, the rejection rate will have an impact on travellers' satisfaction with the SAV service since α is included in the utility calculation. Two pricing rates are tested. S10 and S12 share the same price setting as S1 Base scenario. S11 and S13 share the same price setting as S2 Lower price. Besides, travellers may have different sensitivities to the rejection rate, which is reflected in parameter β_1. A lower value of parameter β_1 is tested in S12 and S13 meaning that travellers can have a lower sensitivity towards the rejection rate.

First, we shall have a look at the trip service rate in different scenarios when the SAV operator is allowed to reject non-profitable trips. Looking at the optimisation results in Table 5, we noticed that S1 Base scenario and S10 Base scenario with rejection yield the same service rate, indicating that the SAV operator did not reject any requests to maintain a high level of service quality despite having the option to decline nonprofitable requests. In S10 Base scenario, β_1 equals 1, and travellers are sensitive to the change in rejection rate. Thus, with this price setting, rejecting trips can decrease demand for SAVs even for profitable requests. However, in S11, when the price rate of using SAVs is lower than in S10 Base scenario, the trip satisfaction rate dropped to 95.44%. This indicates that the SAV operator is willing to accept the loss of revenue caused by decreased travellers' satisfaction and reduced demand in order to save costs by rejecting non-profitable trips. As can be observed in Table 5, the SAV operator earned more profits in S11 Lower price with rejection compared with S2 Lower price while satisfying fewer trips. The cost saving comes from less operational cost, less depreciation cost, and less delay penalty.

When travellers are less sensitive to the service quality level, the trip service rate decreased from 100% (in S10 Base scenario with rejection) to 99.14% (in S12 Base scenario and lower β_1) and from 95.44% (in S11 Lower price with rejection) to 91.6% (in S13

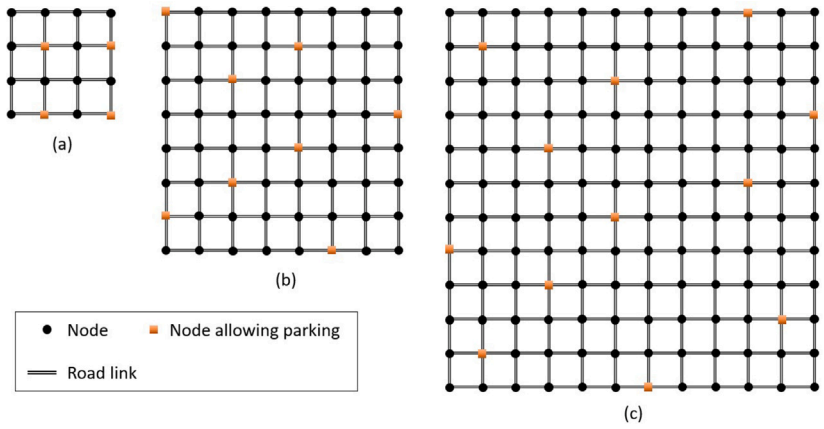


Fig. 10. Illustration of grid networks: (a) small, (b) medium, and (c) large.

Lower price with rejection and lower β_1). It indicates that the SAV operator can increase the profit by rejecting more nonprofitable trips, even if it brings negative impacts on travellers' satisfaction. Although this resulted in a decrease in revenue due to a lower number of satisfied trips, the SAV operator can save on operational costs and reduce delay penalties leading to a higher overall profit.

Rejecting some trips mitigates traffic congestion on the network. As shown in Table 5, the average delay per trip decreases from 1.93 min (in S1 Base scenario and S10 Base scenario with rejection) to 1.68 min (in S12 Base scenario with rejection and lower β_1), and the average delivery time per trip decreases from 16.03 min (in S1 Base scenario and S10 Base scenario with rejection) to 15.88 min (in S12 Base scenario with rejection and lower β_1). The same trend can be found when the price rate of using SAV services is low. The average delay per trip decreases from 3.13 min (in S2 Lower price) to 2.35 min (in S11 Lower price with rejection), then to 1.8 min (in S13 Lower price with rejection and lower β_1), and the average delivery time per trip decreases from 17.93 min (in S2 Lower price) to 17.4 min (in S11 Lower price with rejection), then to 16.65 min (in S13 Lower price with rejection and lower β_1).

In terms of the fleet sizing decisions, we can conclude that these two accept/reject mechanisms do not have a significant impact on the fleet size decisions when the price rate of using SAVs is high. As can be seen in Table 5, the total SAV fleet size in S10 Base scenario with rejection is the same as that in S1 Base scenario, while the total SAV fleet size in S12 Base scenario with rejection and lower β_1 is slightly higher than that in S1 Base scenario and S10 Base scenario with rejection. This indicates that using a bit more vehicles in S12 can save the relocation distance and further release the congestion effect caused by the relocation of SAVs. This part of the savings is greater than the increased depreciation costs of the total fleet which makes it the optimal strategy in S12. However, when the price rate is low, we observe that the fleet size is sensitive to the accept/reject mechanism and parameter β_1. When travellers have a low sensitivity to the service quality level (rejection rate), the SAV operator tends to reject more nonprofitable trips to gain more profits. Thus, a smaller fleet can be deployed as the number of served trips decreases. Our future research will involve further exploration of the spatial impacts of the rejection rate.

6. Scaling analysis: evaluating model performance with various network sizes and demand profiles

Scalability denotes the capability of the proposed methodology to manage an expanding workload, including accommodating larger network sizes and rising demands. Investigating the scalability of our proposed model holds significance due to its nature as a single-level mixed integer programming model that integrates endogenous demand, congestion and accept/reject mechanism, and that is solved using an exact method. Thus, within this section, we present the computational tests conducted to evaluate the performance of the proposed model under various network sizes and demand profiles. These experiments were executed on a desktop computer with an Intel(R) Xeon(R) W-2123 CPU @3.60 GHz, and 32.00 GB RAM. The implementation of the model was accomplished using Python 3.7, and the MILP solver Gurobi 10.0.0 was utilised to solve the optimisation problems.

To evaluate the model's scalability towards the network sizes, we generated three grid networks, each with varying sizes: 16 nodes and 48 directed links, 64 nodes and 224 directed links, and 144 nodes and 528 directed links, as illustrated in Fig. 10. All the links are two-way circulation allowed. For each network, we distributed 4, 8, and 12 parking depots, respectively. The road links in these networks have an equal length of 2 km and an equal capacity of 3200 vehicles/h. In our testing, we set a time step of 2.5 min with the shortest travel time per link at 2.5 min (1 time step) and the longest travel time per link at 10 min (4 time steps).

To investigate the impact of demand variations, we conducted tests with different demand profiles. The total number of trips and the number of groups of trips were modified to simulate the varying passenger demands in different scenarios. The configuration of the tests can be found in Table 7. Trip details such as origin, destination, and departure time were randomly generated to create realistic scenarios. The shortest travel time and the shortest travel distance were calculated using the Dijkstra shortest path algorithm based on the known origin and destination. The latest arrival time for each trip was calculated by doubling the shortest travel time

Table 7
Configurations and computational results (Note: 'N' represents the number of nodes; 'L' represents the number of links; 'P' represents the number of parking depots; 'R' represents the number of trips; 'G' represents the number of group of trips).

Configuration	N	L	P	R	G	Computational time	MIP gap
N16_L48_P4_R1000_G30	16	48	4	1000	30	15 s	0
N16_L48_P4_R2000_G60	16	48	4	2000	60	19 s	0
N16_L48_P4_R3000_G90	16	48	4	3000	90	60 s	0
N64_L224_P8_R1000_G30	64	224	8	1000	30	1112 s≈0.31 h	0
N64_L224_P8_R2000_G60	64	224	8	2000	60	1270 s≈0.35 h	0
N64_L224_P8_R3000_G90	64	224	8	3000	90	7142 s≈1.98 h	0
N144_L528_P12_R1000_G30	144	528	12	1000	30	1984 s≈0.55 h	0
N144_L528_P12_R2000_G60	144	528	12	2000	60	4192 s≈1.16 h	0
N144_L528_P12_R3000_G90	144	528	12	3000	90	9874 s≈2.74 h	0
N144_L528_P12_R6000_G90	144	528	12	6000	90	10 637 s≈2.95 h	0
N144_L528_P12_R9000_G90	144	528	12	9000	90	21 925 s≈6.09 h	0
N144_L528_P12_R6000_G180	144	528	12	6000	180	26 892 s≈7.47 h	0
N144_L528_P12_R9000_G180	144	528	12	9000	180	36 615 s≈10.17 h	0
N144_L528_P12_R12000_G180	144	528	12	12 000	180	>24 h	–

and adding it to the departure time. The optimisation period contains 29 time steps, the same as the case study of the city of Delft in the Netherlands. Furthermore, the remaining parameters remain consistent with the Delft case study and are provided in Table 2 for reference.

The computation times for each of the tests are summarised in Table 7. In the first 9 instances, we increase the number of trips and the number of groups of trips for each network setting. Notably, there is a clear tendency for increased computation time with a higher number of trips and more groups. Then, we compare the instances with the same number of trips and groups, but with increased network size from small (N16_L48_P4) to medium (N64_L224_P8) and then to large (N144_L528_P12). For these increased network sizes, we observe a consistent trend of increasing computation time. To test the computational limits of the proposed model using the current computer, we further intensified the congestion level by enlarging the number of groups and the total trips in the large grid network (N144_L528_P12). As depicted in the last 6 instances of Table 7, this led to a notable increase in computation time. Solving the final instance with 12 000 trips and 180 groups of trips within the large grid network took more than 24 h in computational time. It is important to note that the performance of the proposed model may vary when executed on different computers and utilising different optimisation solvers.

The computational burden of the proposed model arises from the rapid increase in the number of variables and constraints within the time-space network framework. Particularly, the significant rise in the number of integer variables, such as $PF^{r}_{i_{t_1} j_{t_2}}$ and $X_{i_{t_1} j_{t_2}}$, poses challenges for exact methods like branch-and-bound. To further reduce the computational complexity, the following measures can be adopted: (1) employing a rolling-horizon framework to divide the optimisation period into smaller horizons and subsequently resolving the model within each of these horizons; (2) clustering requests based on their spatial and temporal information; (3) developing tailored algorithms to tackle the issue, such as decomposition-based algorithms or meta-heuristics. It is worth mentioning that all these measures come with the potential drawback of losing optimal solutions.

7. Conclusion and future work

In this paper, we propose a non-convex non-linear mathematical programming model to optimise fleet sizing and management decisions of an SAV service while considering traffic congestion and the non-linear demand of multi-class users (according to income). The congestion effect is measured through a dynamically varying travel time with respect to the traffic flow. Travellers' mode choice behaviour is modelled between SAVs and bicycles, assuming that no private cars are allowed in cities, which is captured through an endogenous binary logit model. The two accept/rejection mechanisms (mandatory vs. non-mandatory acceptance) are explored, and the service level is endogenously determined which can affect travellers' willingness to use SAV services. The computational challenge posed by the non-linear and non-convex nature of the model is addressed through reformulation and the use of outer-inner approximation methods combined with a breakpoint generation algorithm to obtain a relaxed version of the original problem. The reformulated model can be solved using state-of-art solvers, such as Gurobi.

A quasi-real case study of Delft, in The Netherlands, was performed and a sensitivity analysis was carried out to demonstrate the performance of the proposed model and provide managerial insights to SAV operators in a promising future scenario. Results indicated that demand for SAVs, supply strategies of SAV operators, and network performance (traffic congestion) are interdependent with each other. Thus, it is crucial to take their interactions into account when managing fleets in an SAV service system. In terms of the fleet sizing strategy, computational results indicated that the initial distribution of the SAV operator's fleet is greatly impacted by factors such as the population's geographical distribution, land use patterns, and residents' travel behaviour. In addition, the fleet sizing decision is significantly influenced by the pricing strategy, unit operating costs of the SAV fleet, network congestion level, and the value of the parameters β_0. When the price rate is low, the fleet sizing decision is also sensitive to the accept/reject mechanism (mandatory vs. non-mandatory acceptance) and the travellers' sensitivity to the service quality level described by parameter β_1. The fleet sizing decision is insensitive to the change in the delay penalty. When the pricing rate of using SAVs is high, the fleet sizing is insensitive to parameter β_1. In addition, a low price of SAV service will attract more users but it may not necessarily bring a

higher profit because of the increased traffic congestion. Besides, bringing fleets with lower operational costs to the system may earn more profits for an SAV operator through operational cost savings, reduction in delay penalties due to the improved traffic congestion, and lower depreciation costs of their fleets as less fleet is needed, despite the fact that SAVs had to take more detours and relocations.

Results indicate that SAV services are more attractive to travellers with a higher VOTT than those with a lower VOTT. Besides, travellers with a high VOTT are more sensitive to variations in trip length compared with the other classes. For long trips, travellers with high VOTT always prefer SAV services. However, for those with lower VOTT, SAV services are only preferred when the price is low. For middle and short trips, bicycles are preferable in most cases unless the price rate is low.

As a direction for future research, we propose the integration of the following aspects into our model: (1) demand and departure time stochasticity; (2) optimising the pricing strategies; (3) worst-case scenarios in robust optimisation; (4) incorporation of ride-sharing mechanisms within the SAV service system; (5) interaction between SAVs and public transit systems.

CRediT authorship contribution statement

Qiaochu Fan: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Visualization. **J. Theresia van Essen:** Conceptualization, Methodology, Validation, Resources, Writing – review & editing, Supervision. **Gonçalo H.A. Correia:** Conceptualization, Methodology, Validation, Resources, Data curation, Writing – review & editing, supervision.

Acknowledgements

This work was supported by Chinese Scholarship Council (CSC).

Appendix. Problem formulation

See Table A.8 for the mathematical notations.

Mixed integer linear program

$$\max \sum_{r\in R} OM^r_{AV} S^r - cf \cdot V - co\left(\sum_{\left(i_{t_1}, j_{t_2}\right)\in G} l_{ij} F_{i_{t_1} j_{t_2}}\right) - cd \sum_{r\in R}\left(\sum_{t\in T} tE^r_t - a^r S^r - st^r S^r\right) \tag{A.1}$$

where

$$OM^r_{AV} = p^0 + sd^r p, \qquad \forall r \in R. \tag{A.2}$$

subject to:

$$V^r_{AV} = -\beta_0(OM^r_{AV} + VOT^r_{AV} T^r_{AV}) - \beta_1(1-\alpha), \qquad \forall r \in R \tag{A.3}$$

$$n^r P^r_{AV} - 0.5 < D^r_{AV} \le n^r P^r_{AV} + 0.5, \qquad \forall r \in R \tag{A.4}$$

$$\sum_{r\in R} D^r_{AV} = \sum_{h=0}^{\mathcal{H}} 2^h \overline{D}_h \tag{A.5}$$

$$Y_h \le \alpha, \qquad \forall h \in \{0, 1, \dots, \mathcal{H}\} \tag{A.6}$$

$$Y_h \le \overline{D}_h, \qquad \forall h \in \{0, 1, \dots, \mathcal{H}\} \tag{A.7}$$

$$Y_h \ge \alpha + \overline{D}_h - 1, \qquad \forall h \in \{0, 1, \dots, \mathcal{H}\} \tag{A.8}$$

$$\sum_{h=0}^{\mathcal{H}} 2^h Y_h = \sum_{r\in R} S^r \tag{A.9}$$

$$S^r \le D^r_{AV}, \qquad \forall r \in R \tag{A.10}$$

$$S^r = \sum_{j_t \mid \left(o^r_{a^r}, j_t\right)\in G} PF^r_{o^r_{a^r} j_t}, \qquad \forall r \in R \tag{A.11}$$

$$S^r = \sum_{t\in T \mid a^r + st^r \le t \le b^r} E^r_t, \qquad \forall r \in R \tag{A.12}$$

Table A.8
Notation.

Notation	Description
Set	
T	$=\{0,1,2,\dots,\mathcal{T}\}$. Set of time instants in the operation period.
N	Set of nodes.
L	Set of road links between nodes in set N.
G	Set of links in the time-space network.
N_P	Set of nodes allowing parking for SAVs with $N_P \subseteq N$.
R	Set of groups of trips, where each group of trips $r \in R$ has the same origin, destination, departure time, and latest arrival time at the destination.
M	Set of travel modes, with the automated vehicles (AV) and bicycles (B) as the two options.
K	$=\{1,2,\dots,k,\dots,\mathcal{K}\}$. Index set of predetermined breakpoints.
Parameters	
Δt	Time step.
l_{ij}	Length of road link $(i,j) \in L$.
Q_{ij}	Capacity of road link $(i,j) \in L$ in vehicles per time step.
$t_{ij}^{\max}$	Maximum travel time by cars on road link $(i,j) \in L$.
$t_{ij}^{\min}$	Minimum travel time by cars on road link $(i,j) \in L$.
$C_{i_{t_1} j_{t_2}}$	Spatial capacity of road link $(i,j) \in L$ in vehicles that fit on the road link from time instant t_1 to t_2, where $(i_{t_1}, j_{t_2}) \in G$.
α	Trip service rate when all the requests have to be accepted, %.
o^r	Origin node for group of trips $r \in R$.
d^r	Destination node for group of trips $r \in R$.
a^r	Departure time for group of trips $r \in R$.
b^r	Latest arrival time for group of trips $r \in R$.
sd^r	Shortest travel distance for group of trips $r \in R$, in kilometres.
st^r	Shortest travel time assuming free-flow speed for group of trips $r \in R$, in time steps.
n^r	Total number of trips for group $r \in R$.
V_B^r	Deterministic systematic component of the utility of bicycles for group of trips $r \in R$.
OM_m^r	Monetary costs of travellers in group $r \in R$ using mode $m \in M$, in euros.
β_0	Parameter converting costs into utility, utility/euro.
β_1	Parameter converting service rate into utility.
VOT_m^r	Travellers' value of travel time in group r using mode $m \in M$, euros/time step.
T_B^r	Travel time of using bicycles for trips in group $r \in R$.
p^0	Initial base fare for using SAVs, euros/trip.
p	Travel distance-related price for using an SAV, euros/km.
co	Unit driving operational cost of an SAV, euros/km.
cd	Penalty for drop-off delay of passengers, euros/time step.
cf	Depreciation cost in one hour for using an SAV, euros/vehicle.
$(u^k, \ln u^k)$	Coordinates of the kth breakpoint.
$\mathcal{M}_r^1$	Big-M parameter, where $r \in R$.
$\mathcal{M}_{ijt_1t_2}^2$	Big-M parameter, where $t_1, t_2 \in T$, if $t_1 < t_2 \le t_1 + t_{ij}^{\max} - t_{ij}^{\min}, (i,j) \in L$.
Decision variables	
V_{AV}^r	Deterministic systematic component of travellers' utility for using an SAV in group $r \in R$.
T_{AV}^r	Longest SAVs travel time for group $r \in R$.
P_{AV}^r	Probability to choose SAVs for the trips in group $r \in R$.
D_{AV}^r	Total number of trips using SAVs in group $r \in R$.
α	Trip service rate when some requests can be rejected.
V	SAV fleet size.
V_i	Initial distribution of SAVs at parking node $i \in N_p$ at the beginning of a day.
S^r	Total number of trips served by SAVs from group r, where $r \in R$.
$PF_{i_{t_1} j_{t_2}}^r$	Passenger flow in the group of trips $r \in R$ served by an SAV in road link (i,j), from time instant t_1 to t_2. Only defined for $(i_{t_1}, j_{t_2}) \in G, a^r \le t_1 < t_2 \le b^r$. If $t_1 = a^r$, then $i = o^r$.
E_t^r	Total number of passengers in group of trips $r \in R$ arriving at time $t \in T$.
$F_{i_{t_1} j_{t_2}}$	Vehicle flow in road link (i,j) from time instant t_1 to t_2, where $(i_{t_1}, j_{t_2}) \in G$. Note that when $t_1 = 0$, $i \in N_P$, meaning that SAVs have to depart from the parking nodes at the beginning of a day.
W_{i_t}	Total number of SAVs parking at node $i \in N_P$ from time instant t to $t+1$, with $t \in T$.
Z_t^r	Binary variable with $r \in R, t \in T$ if $a^r + st^r \le b^r$.
$X_{i_{t_1} j_{t_2}}$	Binary variable which is 1 when any vehicle travels in road link (i,j) from time instant t_1 to t_2, where $(i_{t_1}, j_{t_2}) \in G$, and 0 otherwise.
A_t^r	Binary variable which is 1 when at least one trip in group $r \in R$ arrives at time $t \in T$, and 0 otherwise.
LN_{AV}^r	Auxiliary continuous variable, where $r \in R$.
LN_B^r	Auxiliary continuous variable, where $r \in R$.
λ_r^k	Binary variable indicating whether an interval $[u^k, u^{k+1}]$ is active or not, where $k \in \{1,2,\dots,k,\dots,\mathcal{K}-1\}$, $r \in R$.
θ_r^k	Convex combination coefficient for breakpoint $k \in K$ for group of trips $r \in R$.

(*continued on next page*)

Table A.8 (*continued*).

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

$$E_t^r = \sum_{i_{t_1} \mid \left(i_{t_1}, d_t^r\right) \in G} PF_{i_{t_1} d_t^r}^r, \qquad \forall r \in R, t \in T \tag{A.13}$$

$$\sum_{j_{t_1} \mid \left(d_t^r, j_{t_1}\right) \in G} PF_{d_t^r j_{t_1}}^r = 0, \qquad \forall r \in R, a^r \le t \le b^r \tag{A.14}$$

$$\sum_{i_{t_1} \mid \left(i_{t_1}, o_t^r\right) \in G} PF_{i_{t_1} o_t^r}^r = 0, \qquad \forall r \in R, a^r \le t \le b^r \tag{A.15}$$

$$\sum_{j_{t_0} \mid \left(j_{t_0}, i_{t_1}\right) \in G} PF_{j_{t_0} i_{t_1}}^r = \sum_{j_{t_2} \mid \left(i_{t_1}, j_{t_2}\right) \in G} PF_{i_{t_1} j_{t_2}}^r, \qquad \forall r \in R, a^r < t_1 < b^r, i \in N, i \ne o^r, i \ne d^r \tag{A.16}$$

$$\sum_{r \in R} PF_{i_{t_1} j_{t_2}}^r \le F_{i_{t_1} j_{t_2}}, \qquad \forall \left(i_{t_1}, j_{t_2}\right) \in G \tag{A.17}$$

$$\sum_{j_{t_1} \mid \left(j_{t_1}, i_t\right) \in G, t_1 < t} F_{j_{t_1} i_t} = \sum_{j_{t_2} \mid \left(i_t, j_{t_2}\right) \in G, t < t_2} F_{i_t j_{t_2}}, \qquad \forall i \in N \setminus N_P, 0 < t < \mathcal{T}, \tag{A.18}$$

$$\sum_{j_{t_1} \mid \left(j_{t_1}, i_t\right) \in G, t_1 < t} F_{j_{t_1} i_t} + W_{i_{t-1}} = \sum_{j_{t_2} \mid \left(i_t, j_{t_2}\right) \in G, t < t_2} F_{i_t j_{t_2}} + W_{i_t}, \qquad \forall i \in N_P, 0 < t < \mathcal{T}, \tag{A.19}$$

$$\sum_{j_t \mid (i_0, j_t) \in G} F_{i_0 j_t} + W_{i_0} = V_i, \qquad \forall i \in N_p \tag{A.20}$$

$$\sum_{i \in N_p} V_i = V \tag{A.21}$$

$$\frac{E_t^r}{n^r} \le A_t^r \le E_t^r, \qquad \forall r \in R, a^r + st^r \le t \le b^r \tag{A.22}$$

$$T_{AV}^r \ge A_t^r (t - a^r), \qquad \forall r \in R, a^r + st^r \le t \le b^r \tag{A.23}$$

$$T_{AV}^r \le A_t^r (t - a^r) + (b^r - a^r)(1 - Z_t^r), \qquad \forall r \in R, a^r + st^r \le t \le b^r \tag{A.24}$$

$$\sum_{t \mid a^r + st^r \le t \le b^r} Z_t^r = 1, \qquad \forall r \in R \tag{A.25}$$

$$\sum_{t_2 \mid \left(i_{t_1}, j_{t_2}\right) \in G} X_{i_{t_1} j_{t_2}} \le 1, \qquad \forall (i,j) \in L, t_1 \in T \tag{A.26}$$

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

$$t_1 + \sum_{t \in T} X_{i_{t_1} j_t} \left(t - t_1\right) \le t_2 + \sum_{t \in T} X_{i_{t_2} j_t} (t - t_2) + (t_1 + t_{ij}^{\max} - t_2)\left(1 - \sum_{t \in T} X_{i_{t_2} j_t}\right), \qquad \forall (i,j) \in L, \quad t_1 < t_2 \le t_1 + t_{ij}^{\max} - t_{ij}^{\min} \tag{A.28}$$

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

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

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

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

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

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

$$\theta_r^1 \le \lambda_r^1, \qquad \forall r \in R \tag{A.35}$$

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

$$\theta_r^{\mathcal{K}} \le \lambda_r^{\mathcal{K}-1}, \qquad \forall r \in R \tag{A.37}$$

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

$$LN_B^r \ge \sum_{k=1}^{\mathcal{K}} \overline{\theta}_r^k \ln u^k, \qquad \forall r \in R \tag{A.39}$$

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

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

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

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

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

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

$$0 \le \alpha \le 1 \tag{A.46}$$

$$V_{AV}^r \ge 0, \qquad \forall r \in R \tag{A.47}$$

$$T_{AV}^r \in \mathbb{N}^0, \qquad \forall r \in R \tag{A.48}$$

$$P_{AV}^r \ge 0, \qquad \forall r \in R \tag{A.49}$$

$$D_{AV}^r \in \mathbb{N}^0, \qquad \forall r \in R \tag{A.50}$$

$$\overline{D}_h \in \{0, 1\}, \qquad \forall h \in \{0, 1, \dots, \mathcal{H}\} \tag{A.51}$$

$$Y_h \ge 0, \qquad \forall h \in \{0, 1, \dots, \mathcal{H}\} \tag{A.52}$$

$$V \in \mathbb{N}^0 \tag{A.53}$$

$$V_i \in \mathbb{N}^0, \qquad \forall i \in N_P \tag{A.54}$$

$$S^r \in \mathbb{N}^0, \qquad \forall r \in R \tag{A.55}$$

$$E_t^r \in \mathbb{N}^0, \qquad \forall r \in R, t \in T \tag{A.56}$$

$$PF_{i_{t_1} j_{t_2}}^r \in \mathbb{N}^0, \qquad \forall r \in R, (i_{t_1}, j_{t_2}) \in G \tag{A.57}$$

$$F_{i_{t_1} j_{t_2}} \ge 0, \qquad \forall (i_{t_1}, j_{t_2}) \in G \tag{A.58}$$

$$W_{i_t} \geq 0, \qquad \forall i \in N_P, t \in T \tag{A.59}$$

$$Z_t^r \in \{0,1\}, \qquad \forall r \in R, t \in T, a^r + st^r \leq b^r \tag{A.60}$$

$$X_{i_{t_1} j_{t_2}} \in \{0,1\}, \qquad \forall (i_{t_1}, j_{t_2}) \in G \tag{A.61}$$

$$A_t^r \in \{0,1\}, \qquad \forall r \in R, t \in T, a^r + st^r \leq t \leq b^r \tag{A.62}$$

$$LN_{AV}^r \geq 0, \qquad \forall r \in R \tag{A.63}$$

$$LN_B^r \geq 0, \qquad \forall r \in R \tag{A.64}$$

$$\theta_r^k \geq 0, \qquad \forall r \in R, k \in K \tag{A.65}$$

$$\lambda_r^k \in \{0,1\}, \qquad \forall r \in R, k \in K \tag{A.66}$$

$$\overline{\theta}_r^k \geq 0, \qquad \forall r \in R, k \in K \tag{A.67}$$

$$\overline{\lambda}_r^k \in \{0,1\}, \qquad \forall r \in R, k \in K \tag{A.68}$$

References

Ashkrof, P., Homem de Almeida Correia, G., Cats, O., van Arem, B., 2019. Impact of automated vehicles on travel mode preference for different trip purposes and distances. Transp. Res. Rec. 2673 (5), 607–616.

Atasoy, B., Salani, M., Bierlaire, M., 2014. An integrated airline scheduling, fleeting, and pricing model for a monopolized market. Comput.-Aided Civ. Infrastruct. Eng. 29 (2), 76–90.

Azadeh, S.S., van der Zee, J., Wagenvoort, M., 2022. Choice-driven service network design for an integrated fixed line and demand responsive mobility system. Transp. Res. A 166, 557–574.

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.

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.

Cai, Y., Chen, J., Lei, D., Yu, J., et al., 2022. The integration of multimodal networks: The generalized modal split and collaborative optimization of transportation hubs. J. Adv. Transp. 2022.

Chung, J.-H., Hwang, K.Y., Bae, Y.K., 2012. The loss of road capacity and self-compliance: Lessons from the cheonggyecheon stream restoration. Transp. Policy 21, 165–178.

Correia, G.H., Looff, E., Van Cranenburgh, S., Snelder, M., Van Arem, B., 2019. On the impact of vehicle automation on the value of travel time while performing work and leisure activities in a car: Theoretical insights and results from a stated preference survey. Transp. Res. A 119, 359–382.

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 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.

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

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

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

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

Gurumurthy, K.M., de Souza, F., Enam, A., Auld, J., 2020. Integrating supply and demand perspectives for a large-scale simulation of shared autonomous vehicles. Transp. Res. Rec. 2674 (7), 181–192.

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 123, 102974.

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 142, 102081.

Huang, Y., Kockelman, K.M., 2020. Electric vehicle charging station locations: Elastic demand, station congestion, and network equilibrium. Transp. Res. D 78, 102179.

Hulse, L.M., 2023. Pedestrians' perceived vulnerability and observed behaviours relating to crossing and passing interactions with autonomous vehicles. Transp. Res. F 93, 34–54.

Hyland, M.F., Mahmassani, H.S., 2017. Taxonomy of shared autonomous vehicle fleet management problems to inform future transportation mobility. Transp. Res. Rec. 2653 (1), 26–34.

Joksimovic, D., Bliemer, M.C., Bovy, P.H., 2005. Optimal toll design problem in dynamic traffic networks with joint route and departure time choice. Transp. Res. Rec. 1923 (1), 61–72.

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. Optimization of Urban Transportation Service Networks, Transp. Res. B Optimization of Urban Transportation Service Networks, 81.461–482,
Kaufman, D.E., Nonis, J., Smith, R.L., 1998. A mixed integer linear programming model for dynamic route guidance. Transp. Res. B 32 (6), 431–440.
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 129, 155–169.
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 112, 260–281.
Liang, X., Homem Correia, G., Van Arem, B., 2018. Applying a model for trip assignment and dynamic routing of automated taxis with congestion: system performance in the city of Delft, the Netherlands. Transp. Res. Rec. 2672 (8), 588–598.
Liu, H., Wang, D.Z., 2015. Global optimization method for network design problem with stochastic user equilibrium. Transp. Res. B 72, 20–39.
Lou, Y., Yin, Y., Laval, J.A., 2011. Optimal dynamic pricing strategies for high-occupancy/toll lanes. Transp. Res. C 19 (1), 64–74.
Lu, R., Correia, G.H.d.A., 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.
Madigan, R., Nordhoff, S., Fox, C., Amini, R.E., Louw, T., Wilbrink, M., Schieben, A., Merat, N., 2019. Understanding interactions between automated road transport systems and other road users: A video analysis. Transp. Res. F 66, 196–213.
Nieuwenhuijsen, M.J., Khreis, H., 2016. Car free cities: Pathway to healthy urban living. Environ. Int. 94, 251–262.
Oh, S., Seshadri, R., Azevedo, C.L., Kumar, N., Basak, K., Ben-Akiva, M., 2020. Assessing the impacts of automated mobility-on-demand through agent-based simulation: A study of Singapore. Transp. Res. A 138, 367–388.
Paneque, M.P., Bierlaire, M., Gendron, B., Azadeh, S.S., 2021. Integrating advanced discrete choice models in mixed integer linear optimization. Transp. Res. B 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., Hyland, M.F., Mahmassani, H.S., Verbas, I.Ö., 2020. Joint design of multimodal transit networks and shared autonomous mobility fleets. Transp. Res. C 113, 2–20.
Ren, X., Chow, J.Y., 2022. A random-utility-consistent machine learning method to estimate agents' joint activity scheduling choice from a ubiquitous data set. Transp. Res. B 166, 396–418.
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.
Tennøy, A., 2010. Why we fail to reduce urban road traffic volumes: Does it matter how planners frame the problem? Transp. Policy 17 (4), 216–223.
Tian, J., Jia, H., Wang, G., Lin, Y., Wu, R., Lv, A., 2022. A long-term shared autonomous vehicle system design problem considering relocation and pricing. J. Adv. Transp. 2022.
Uber, 2023. Uber. URL https://www.uber.com/.
Van Essen, J.T., Correia, G.H., 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.
Vlakveld, W., van der Kint, S., Hagenzieker, M.P., 2020. Cyclists' intentions to yield for automated cars at intersections when they have right of way: Results of an experiment using high-quality video animations. Transp. Res. F 71, 288–307.
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. Physica A 605, 128033.
Wang, D.Z., Liu, H., Szeto, W., 2015. A novel discrete network design problem formulation and its global optimization solution algorithm. Transp. Res. E 79, 213–230.
Wang, D.Z., Lo, H.K., 2010. Global optimum of the linearized network design problem with equilibrium flows. Transp. Res. B 44 (4), 482–492.
Wang, S., Mo, B., Zhao, J., 2020. Deep neural networks for choice analysis: Architecture design with alternative-specific utility functions. Transp. Res. C 112, 234–251.
Wei, K., Vaze, V., Jacquillat, A., 2022. Transit planning optimization under ride-hailing competition and traffic congestion. Transp. Sci. 56 (3), 725–749.
Xu, X., Chen, A., Jansuwan, S., Yang, C., Ryu, S., 2018b. Transportation network redundancy: Complementary measures and computational methods. Transp. Res. B 114, 68–85.
Xu, M., Meng, Q., 2020. Optimal deployment of charging stations considering path deviation and nonlinear elastic demand. Transp. Res. B 135, 120–142.
Xu, M., Meng, Q., Liu, Z., 2018a. Electric vehicle fleet size and trip pricing for one-way carsharing services considering vehicle relocation and personnel assignment. Transp. Res. B 111, 60–82.
Yang, S., Wu, J., Sun, H., Qu, Y., Wang, D.Z.W., 2022. Integrated optimization of pricing and relocation in the competitive carsharing market: A multi-leader-follower game model. Transp. Res. C 138, 103613.
Ye, J., Jiang, Y., Chen, J., Liu, Z., Guo, R., 2021. Joint optimisation of transfer location and capacity for a capacitated multimodal transport network with elastic demand: a bi-level programming model and paradoxes. Transp. Res. E 156, 102540.
You, L., He, J., Zhao, J., Xie, J., 2022. A federated mixed logit model for personal mobility service in autonomous transportation systems. Systems 10 (4), 117.
Zhang, W., Guhathakurta, S., 2017. Parking spaces in the age of shared autonomous vehicles: How much parking will we need and where? Transp. Res. Rec. 2651 (1), 80–91.