Simulation-based assessment of operational ridesharing strategies for shared autonomous vehicles in large-scale networks

Zhou *et al. European Transport Research Review* (2026) 18:28
https://doi.org/10.1186/s12544-026-00787-4

European Transport Research Review

ORIGINAL PAPER Open Access

Simulation-based assessment of operational ridesharing strategies for shared autonomous vehicles in large-scale networks

Ze Zhou[1,2*], Serio Agriesti[2,3,4], Claudio Roncoli[2,5], Lampros Yfantis[6], Jordi Casas[7] and Bat-hen Nahmias-Biran[8]

Abstract
Dynamic ridesharing (DRS) is envisioned as a sustainable mobility solution, with shared automated vehicles (SAVs) expected to facilitate DRS operations in the near future. Existing DRS-SAV-related literature typically focuses either on optimising fleet operations or simulating SAVs employing heuristic DRS methods, which may lead to an incorrect estimation of their true impacts. In this study, we design a time-based framework that enables communication between the DRS service provider and a mesoscopic traffic simulator. We incorporate individual socio-demographic features to generate future demand for SAVs and conduct two sets of simulation experiments within a detailed, large-scale urban network. First, we compare advanced DRS operation strategies during the morning peak. Then, we test the most effective strategy over an extended simulation period to replicate full business day operations. The results indicate that optimisation-based matching combined with fairness-aware pricing outperforms heuristic strategies in improving ridesharing performance and alleviating network congestion. Furthermore, the extended simulations reveal that profit-oriented objectives, which are common among transport network companies, result in significantly more empty VKT. To encourage the adoption of ridesharing, transport authorities should consider penalising (e.g., taxing) solo trips or subsidising shared trips.

Keywords Dynamic ridesharing, Shared autonomous vehicle, Fleet operation, Pricing, Congestion

1 Introduction

Excessive private vehicles, together with the increasing need for mobility, have been constantly challenging the existing transport systems, particularly in densely populated cities. Transport network companies, such as Uber, Lyft, and Didi, which provide internet-based on-demand mobility services, are envisioned to facilitate a shift from car ownership to car usage, possibly enabling a more sustainable utilisation of the vehicle fleets. Furthermore, dynamic ridesharing (DRS), referring to using one vehicle to serve multiple travellers whose trip itineraries overlap to a certain degree, with trip information being disclosed over time [1], is expected to increase vehicle occupancy. By potentially (slightly) sacrificing time and comfort, DRS passengers benefit from sharing the travel fare with their

*Correspondence:
Ze Zhou
ze.zhou@chalmers.se
[1]Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden
[2]Department of Built Environment, Aalto University, Espoo, Finland
[3]FinEst Centre for Smart Cities, Tallinn University of Technology, Tallinn, Estonia
[4]Department of Technology, Management and Economics, Technical University of Denmark, Lyngby, Denmark
[5]Mobility and Industrial Management, KU Leuven, Leuven, Belgium
[6]Ertico ITS Europe, Bruxelles, Belgium
[7]Department of Economics and Business, University of Vic - Central University of Catalonia, Vic, Spain
[8]School of Mechanical Engineering, Tel Aviv University, Tel Aviv, Israel

co-riders, thus the combined convenience and cost-effectiveness of on-demand services, including ride-hailing and DRS, are expected to attract travellers [2]. Nonetheless, the operational challenges posed by DRS hinder its widespread adoption [3]. Automated vehicle (AV) technology is seen as complementary to DRS service, as AVs are expected to accommodate more conveniently to frequent route and assignment changes. Moreover, shared AVs (SAVs) are envisioned to benefit travellers [4], as well as the transport systems and the environment [5] by reducing the negative externalities of owning and driving a personal car.

Most studies about SAVs and DRS either emphasise fleet operation or traffic simulation. The operation-focused studies include the design of novel matching [6, 7] and pricing [8] strategies, which are then tested using simulation tools that consider simplistic traffic conditions, such as static or historical travel time. On the other hand, large-scale simulation-based studies [5, 9] commonly employ simple heuristic operation methods to evaluate the impact of SAVs on traffic. However, such methods may underestimate the effect of DRS, as demonstrated by [10], where comparing DRS-SAV to personal non-autonomous vehicles without sharing showed no significant difference. This study evaluates the system-level performance of optimisation-based matching and fairness-aware pricing strategies for demand-responsive SAV services, and compares them against commonly used heuristic approaches under realistic congestion and traveller response dynamics.

This paper builds on the work recently published by [8], where an algorithm designed for fair pricing and optimisation-based vehicle-traveller matching is presented, which, to the best of our knowledge, is the first work in literature coupling fair pricing and fleet matching optimisation. However, their work relied on stringent assumptions, the most significant of which was the use of static travel times within the network. In this work, we overcome this limitation by testing the proposed algorithm in a mesoscopic traffic assignment tool, capable of producing dynamic travel times and replicating congestion propagation. By doing so, we prove that the developed algorithms outperform state-of-the-art ones even in settings where queue propagation is accounted for, further proving the effectiveness of the proposed solutions. As it is widely acknowledged that on-demand SAVs will have an impact on the overall traffic efficiency [11, 12], to benchmark the proposed solution through dynamic traffic assignment allows to validate its performance in a realistic setting. The results provide an important glimpse into what are the implications of an on-demand SAV service for a medium sized European city, when state-of-the-art methodologies are combined to properly frame the multiple dimensions of the problem (i.e. matching, pricing and dynamic traffic simulation). It does so by evaluating fleet and service performance in a realistic setting but also by assessing the impacts on the overall network congestion.

2 Literature review

2.1 DRS operations: matching and pricing

The DRS problem is a variant of the classic vehicle-routing problem and dynamic pickup and delivery problem [13, 14]. In [15] an optimisation-based method is introduced to assess the potential of DRS using real travel data in Atlanta, finding that using optimisation-based methods instead of simple heuristic rules could substantially improve the performance of DRS. Early studies [15–17] relied on heuristic methods or integer linear program (ILP) formulation [18]. Due to computational limitations, these ILP models do not apply to large-scale scenarios. In [6] this research gap was filled by proposing an efficient algorithm, which has been validated using the New York City taxi dataset. Subsequently, a computationally efficient method was proposed in [19] by formulating the DRS matching into a classic linear assignment problem, which is even up to four times faster than the one in [6] while achieving similar service quality. In [20] an extended linear assignment problem was formulated to account for endogenous congestion caused by the ridesharing fleet.

Nonetheless, all these DRS studies do not deal with ridesharing pricing and make strong assumptions that travellers would accept any ride the platform provides. The authors of [21] proposed a fair fare split method such that every co-rider is fully satisfied with her/his trip fare and assignment. In [22], the authors investigated how to fairly split the ridesharing trip cost among co-riders such that the optimal solution is also an equilibrium. Furthermore, a dynamic discount strategy was designed in [23] to incentivise travellers to choose ridesharing trips. In [24] a novel taxi ridesharing strategy was developed accounting for both online and offline travellers. Besides, they introduced a ridesharing payment model to facilitate the calculation of ridesharing fares. The authors of [25] introduced an operational framework of dial-a-ride problem considering both pricing and customer choice. Additionally, in [8] a scalable fairness-aware pricing method was proposed to charge ridesharing passengers based on their detour and shared trip distance, where fairness is defined from the traveller's perspective. However, optimisation-focused studies usually lack realistic traffic flow models or even assume constant speed. Nevertheless, ignoring the endogenous congestion effects and travellers' mode choices may yield over-optimistic results [5].

2.2 SAV + DRS simulation

Simulation-focused studies attempt to use state-of-the-art traffic simulators to evaluate the DRS impact on operation performance metrics and network congestion. The authors of [10] argued that previous studies might exaggerate the impact of SAV due to the unrealistic congestion models and travel demands they used and, to address this issue, proposed an event-based framework for simulating SAVs using a traffic simulator based on the cell transmission model. A stream of works utilised the agent-based simulator MATsim, which employs a queue-based approach to model traffic dynamics [26]; The work in [27] showed that DRS is critical to reduce the vehicle-miles travelled, thereby traffic congestion, demonstrating the benefits brought by DRS-enabled SAVs; in [28], the performance of traditional taxis was compared with DRS-enabled SAVs, indicating that DRS can significantly increase vehicle occupancy and reduce carbon emissions; in [5], the benefits and costs of DRS-enabled SAVs in Austin, Texas, were quantified demonstrating that, in the presence of both private AVs and SAVs, DRS is beneficial to the transport system when road pricing is enforced during peak hours. The authors of [29] simulated SAVs in a multimodal dynamic demand system to evaluate the performance of the SAV fleet, concluding that ridesharing and vehicle rebalancing can greatly improve the system performance. More focus on the behavioural component has been dedicated in [30–32], which exploited the behavioural activity-based model SimMobility MT to model the utilities of new automated services and how these compare to current modal alternatives. Another example is the work in [33], where different scenarios for automated mobility on-demand have been assessed by exploiting Aimsun (for the supply) and SimMobility MT (for the demand). SimMobility MT is also used in this work to identify the initial, potential demand of the SAV service. More details about the SimMobility MT can be found in [31].

As mentioned above, there is a gap between operation-focused and simulation-focused studies, while only a few studies have tried to fill this gap. This is primarily due to the high computational burden of repeatedly solving optimisation problems under dynamic traffic conditions, the complexity of interfacing optimisation algorithms with traffic simulators, and the difficulty of ensuring real-time consistency between vehicle routing, congestion propagation, and traveller responses. As a result, most existing studies either rely on optimisation models with simplified or static travel times, or on traffic simulations employing heuristic operational rules. A notable exception is the work of [34], which integrates a Mobility-on-Demand simulation framework with Aimsun to explore how dynamic and stochastic travel times affect fleet metrics. Results show that unreliable pickup time and inaccurate travel time estimation, stemming from static or deterministic travel time assumptions, deteriorate the performance of MoD services. However, they focus on dynamic travel time without modelling pricing and traveller's behaviour. In both [5] and [4] DRS-enabled SAV were simulated and account for traveller's mode choice. However, only simple rule-based ridesharing operation methods were adopted.

2.3 Contribution

In this paper, we employ the recently developed matching and pricing method proposed in [8] and overcome its most restricting hypotheses. Building on its computationally efficient optimisation structure, the proposed simulation framework demonstrates how optimisation-based matching and pricing can be effectively integrated with a mesoscopic traffic simulator for large-scale SAV operations. The main differences are summarised as follows:

1. This study employs state-of-the-art traffic modelling software to simulate both ridesharing and private vehicles within a detailed large-scale urban network.
2. The ridesharing operator can dynamically compute the shortest paths and assignment plans for vehicles based on real-time traffic conditions.
3. Travellers make mode choice decisions based on real-time travel times and trip fares.

The remainder of the paper is structured as follows: Sect. 3 describes the simulation framework and ridesharing algorithms; in Sect. 4, two simulation experiments are conducted for the city of Tallinn, Estonia, to evaluate different operation strategies and the contribution of different components; Sect. 5 offers an analysis and discussion of the obtained results; finally, Sect. 6 concludes the paper and suggests directions for future research.

3 Methodology

To test the developed solutions within a dynamic traffic simulation environment, while still accounting for an elastic demand, we propose a time-based framework for implementing optimisation-based DRS-SAV in existing traffic simulators. Compared to the event-based framework, where each request triggers a vehicle assignment action, the time-based framework allows the system to group all the requests during a short interval and then optimise the vehicle assignment plan collectively. We then use the framework to test the matching and pricing algorithms from [8]. Travellers' mode choices are modelled based on real-time travel times and the ridesharing service performance.

3.1 Simulation framework

In this section, we first introduce a time-based simulation framework that integrates the simulation tool Aimsun Ride [35] with the newly developed service operator. Next, we briefly discuss how the operator dispatches vehicles and charges users. Finally, we explain how users choose their travel mode based on the ride offers.

To precisely depict the network traffic, we employ the commercial software Aimsun to simulate multimodal large-scale urban traffic. Particularly, Aimsun Ride is a simulation platform for planning towards new and primarily demand-responsive mobility in urban environments. It is an agent-based demand-supply interaction framework for multimodal and multi-operator fleet-based service systems, coupled with the Aimsun Next's multi-class mesoscopic traffic simulator [36]. It should be noted that an operator in this context refers to a service provider or fleet manager. In terms of the traffic model, the well-known car-following model based on the Gipps model [37, 38] is implemented in Aimsun Next.

In this paper, we develop a time-based simulation framework that enables real-time communication between a ridesharing platform and a traffic simulator. Fig. 1 illustrates the time-based communication between Aimsun Ride and the Operator modules. Three input files include requests, fleet, and the network. To begin with, all requests are passed to the service operator, while Aimsun Ride is called by the service operator at a fixed time interval Δt. A callback is sent from Aimsun Ride to the Service Operator asking for an action plan every Δt. At each time step, all requests arriving within the interval $[t, t + \Delta t]$ are jointly considered by the operator, which solves the matching and pricing problem and updates vehicle schedules accordingly. Specifically, the Service Operator is responsible for determining the vehicle assignment (including idle vehicle assignment) and pricing plans, by performing a series of algorithms proposed in [6]. Then the traveller can either accept or reject the ride offer based on the price and trip information. Note that pricing and matching are two critical components in DRS operation. We later tested different strategies in the experiments. Once the offer is confirmed, the operator communicates the final schedule for each vehicle-request pair to Aimsun Ride. The simulator implements the operational decisions and simulates the traffic in one time step. The changed vehicle status and traffic conditions are available to the service operator in real-time, facilitating the next round of assignment. These steps are performed iteratively until the end of the simulation.

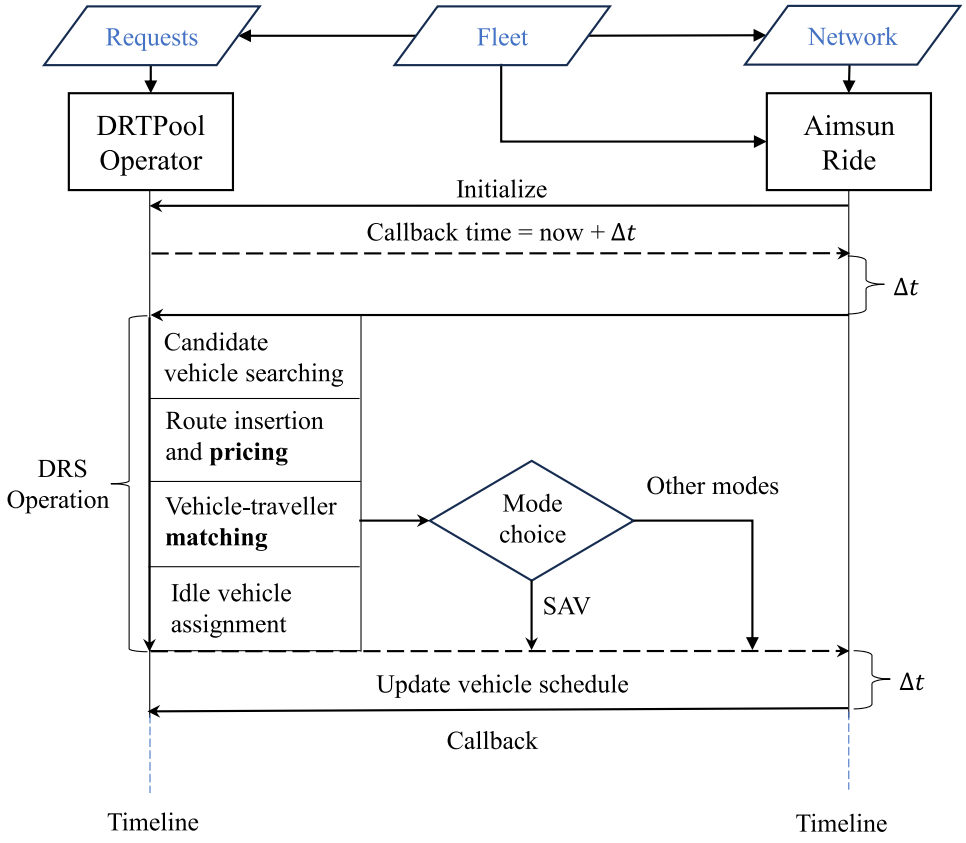


Fig. 1 Time-based communication sequence diagrams between Service Operator and Aimsun Ride

3.2 Ridesharing operation algorithm

In this study, ridesharing operation involves vehicle-traveller matching, pricing, and idle vehicle dispatching.

In particular, the fleet operator batches the requests within a fixed time interval and searches candidate vehicles for each request based on the pickup time. In cases where no available vehicle is found, for instance when all the nearby vehicles violate the pickup time window, we assume the traveller would either wait until being assigned to a vehicle or leave the system if the maximum waiting time constraint is violated. In the latter case, a private vehicle is assumed to serve the trip. Otherwise, the operator calculates the trip fare according to the vehicle schedule associated with each potential vehicle-traveller pair. For solo trips, we adopt the conventional taxi fare structure to calculate the fare f_r^{solo}:

$$f_r^{\text{solo}} = f_0 + f_t t_r + f_d d_r, \tag{1}$$

where f_0, f_t, and f_d denote base, time, and distance fare rates, respectively, while t_r and d_r are the solo trip time and distance, respectively. For a ridesharing trip, the new traveller's origin-destination is inserted into the vehicle's schedule to minimise total travel time. While ridesharing, the trip fare f_r^{share} is calculated via:

$$f_r^{\text{share}} = \Phi_r \cdot \hat{f}_r, \tag{2}$$

where $\hat{f}_r$ denotes the previous price before route insertion and the multiplicative discount function $0 < \Phi(\cdot) \leq 1$ accounts for the traveller's detour rate, the number of co-riders and shared trip distance. Note that $\hat{f}_r = f_r^{\text{solo}}$ for every new request r. Furthermore, the trip fares for those assigned in the earlier stages are changed if the trip is shared with the new traveller.

Specifically, when ridesharing is considered, the vehicle schedule is updated and a discount factor $\Phi_r = \Phi_r(\tau_r^{\text{d}}, \zeta_r^{\text{g}})$ is computed based on two indices: the relative detour index τ_r^{d}, capturing the additional travel experienced by passenger r; and the ridesharing index ζ_r^{g}, which quantifies the extent to which the passenger's trip distance and travel time are shared with other riders. By construction, $\Phi_r(\tau_r^{\text{d}}, \zeta_r^{\text{g}}) \in (0, 1]$, the ridesharing fare is always lower than the corresponding solo fare, and a shared trip is considered valid only if the updated price is strictly lower than the pre-sharing price. Consequently, no rider pays more after sharing. If this condition is violated due to numerical or scheduling effects, the offer is classified as unfair and automatically rejected. Appendix A reports the formulation of the ridesharing discount function. Readers can refer to [8] for a detailed explanation.

The operator is tasked with optimising the assignment plan, determining which vehicle should serve which traveller's request. In this paper, we compare two matching strategies: greedy matching, which is the most popular method in simulation-based studies, and optimisation-based matching. In the former, a traveller is paired with a candidate vehicle with the minimum travel time to serve the request. In the latter, an integer linear optimisation problem for ridesharing trips is formulated to determine the optimal vehicle-traveller matching.

Based on Eqs. (1)-(2), we define the trip profit as

$$\pi_{P,v} = \begin{cases} \delta \cdot (f_r^{\text{solo}} - c_{P,v}), & \text{P} = 1, \text{vehicle is idle} \\ \left(\sum_{r \in P} f_r^{\text{share}} - c_{P,v}\right) - \left(\sum_{r \in \hat{P}} \hat{f}_r - c_{\hat{P},v}\right), & \text{P} > 1, \text{otherwise} \end{cases} \tag{3}$$

where δ is a parameter designed to scale the solo trip profit when the vehicle is idle; $c_{P,v}$ denotes the trip cost, such as energy consumption, for vehicle v to serve traveller p; P denotes the set including all assigned travellers except those who have reached their destination; $\hat{P}$ represents the same set before the current traveller joins; $C_{\hat{P},v}$ and $C_{P,v}$ indicate the total trip cost for serving $\hat{P}$ and P, respectively. Note that, when the vehicle is not idle, the trip profit $\pi_{P,v}$ for vehicle v is defined as the marginal profit incurred each time a new traveller is added to its schedule.

The optimisation problem is formulated as

$$\max_x \sum_r \sum_{v \in V_r} \pi_{P,v} x_{r,v} - \gamma(|R| - \sum_r \sum_{v \in V_r} x_{r,v}) \tag{4}$$

subject to

$$\sum_{v \in V_r} x_{r,v} \leq 1, \quad \forall r \in R \tag{5}$$

$$\sum_r x_{r,v} \leq 1, \quad \forall v \in V_r \tag{6}$$

$$x_{r,v} \in \{0, 1\}, \tag{7}$$

where V_r represents the candidate vehicle set for traveller r; $x_{r,v}$ is a binary variable that equals 1 if vehicle v is assigned to serve traveller r and 0 otherwise; the parameter γ is the weight for the penalty term accounting for unserved travellers; and $|R|$ denotes the number of travellers that wait to be served; therefore, the objective (4) aims to maximise the difference between total profit and unserved penalty. The vehicle schedule and trip fare are determined by solving the above optimisation problem. Next, the operator proceeds to finalise the vehicle-traveller matching and delivers the trip information to the traveller.

Upon receiving the offer, the traveller, based on the received price and travel time, can either accept or decline the ride. Notably, the model accounts for two

types of rejection, either because of unfair pricing or inappropriate trip arrangements, such as long travel time. When the offer is unfair, for example, the ridesharing price is even higher than the price before sharing, then the traveller is assumed to leave the system directly and use a private vehicle instead. In this context, the unfair price refers to the ridesharing price that disregards the traveller's detour and shared distance, resulting in a higher overall cost. On the other hand, if a trip is not preferred due to longer waiting or detour times, we assume travellers would stay in the system and wait for the next round of assignment. It is worth mentioning that our method allows for the modelling of both human-driven and autonomous private vehicles.

Finally, due to supply-demand imbalances, idle vehicles may remain in a low-demand region and are unable to serve future requests. To address this, we implement a reactive, demand-driven idle vehicle repositioning algorithm from [6] to assign idle SAVs to unmatched travellers. Specifically, idle SAVs are not proactively relocated based on demand forecasts. Instead, repositioning is triggered only when there exists an imbalance between waiting requests and idle vehicles. In such cases, the operator solves a small assignment problem that matches idle vehicles to currently unmatched requests by minimizing travel time to request origins, subject to one-to-one assignment constraints. If a request accepts the service, the vehicle proceeds as a pickup; otherwise, the vehicle is repositioned toward the request's origin area to improve near-term availability. Note that idle vehicle assignment is essential for maintaining a continuous and effective on-demand service, particularly under conditions of supply–demand imbalance. Nonetheless, we emphasize that vehicle repositioning is not a primary focus of this study; rather, it is included for completeness and to enable system-level performance evaluation. Our analysis instead centers on how the proposed pricing and matching strategies influence eventual repositioning demand and empty VKT.

3.3 Mode choice

Travellers submit their requests, providing travel origin and destination, maximum waiting time and acceptable detour time. Note that we consider travellers with heterogeneous waiting and detour tolerances. We utilise the following formulations to calculate the utility of different travel modes:

$$f_{\mathrm{PV}} = f^0_{\mathrm{PV}} + f^v_{\mathrm{PV}} \cdot d_{\mathrm{PV}} \tag{8}$$

$$u_{\mathrm{PV}} = \beta_{\mathrm{PV}} + \beta_t \cdot t_{\mathrm{PV}} + \beta_f \cdot f_{\mathrm{PV}} \tag{9}$$

$$u_{\mathrm{SAV}} = \beta_{\mathrm{SAV}} + \beta_t \cdot t_{\mathrm{SAV}} + \beta_f \cdot f_{\mathrm{SAV}}, \tag{10}$$

where f_{PV} represents the private trip cost and PV is the acronym for private vehicles; f^0_{PV} denotes the fixed cost per trip, indicating the vehicle depreciation rate per trip; f^v_{PV} is variable costs per kilometre, representing fuel consumption expense; d_{PV} is the private trip distance; u_{PV} and u_{SAV} are the travel utilities of private and SAV trips, respectively; β_{PV} and β_{SAV} are model-specific constants to reflect the inherent trip qualities of private vehicles and SAV, respectively; β_t and β_f represent the marginal disutility of travel time and trip fare; t_{PV}, t_{SAV}, and f_{SAV} are travel time and trip fare for private and SAV trip respectively. Note that f_{SAV} can be calculated via (1) and (2) depending on whether the trip is shared or not. Although we do not explicitly model public transit in this study, it can be easily incorporated using a specific formulation. Travellers either accept or decline the SAV offer based on the trip utility, which is modelled here via a logit model [39]. For the travellers who were assigned before the current batch and shared with new travellers (requests), we assume they would automatically accept the updated price only if it is lower than the original price. The new traveller refers to the travellers that need to be assigned in the current batch, rather than those assigned in the previous steps.

Note that public transport is not explicitly modelled in the real-time mode choice; however, as described later in Sect. 4.2, its role is partially captured upstream through the activity-based demand generation, where it constitutes a substantial competing mode. We focus on short-term competition between SAVs and private vehicles to isolate the operational and congestion effects of different ridesharing strategies. Public transport can be incorporated by adding a transit utility, based on travel time, waiting time, and fare, to the logit model using schedule- or simulation-based transit outputs.

4 Simulation settings

4.1 Basic settings

We use the City of Tallinn, Estonia, as our case study. For our experiments, we employ a network that covers an area of 240 km^2, including 33,000 sections and 15,600 nodes. We select Tallinn as the case study area due to the availability of a detailed, calibrated large-scale traffic model and an activity-based demand model. As a medium-sized European city with recurrent congestion and diverse travel patterns, Tallinn provides a realistic and representative testbed for assessing the operational and network-level impacts of shared automated vehicle services. The model has been built in a way suitable to reproduce the door-to-door service provided by SAVs, as the precision with which origins and destinations are framed equals the area of each of the 610 centroids (500x500 meters, see Fig. 2). Each request is therefore served within walking distance from both their origin and their destination.

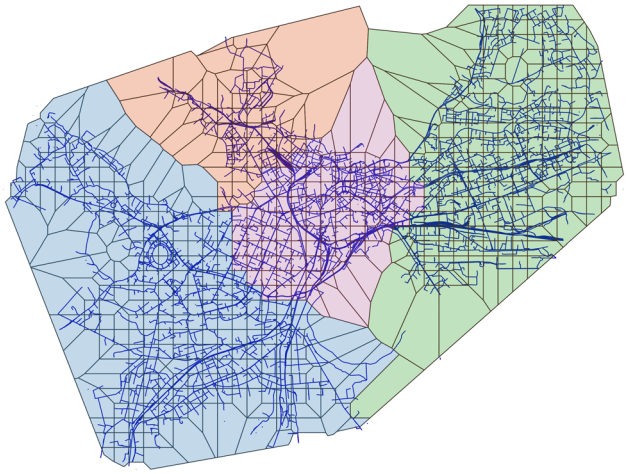


Fig. 2 Centroids and the traffic network - shades of colours based on the area of the city: purple - centre, blue - western, orange - northern, green - eastern

In (3) and (4), the solo-trip scaling factor is fixed at $\delta = 0.7$, while the unserved-request penalty is set to $\gamma = 2€$ per request. The parameter δ mildly penalises solo trips to encourage ridesharing while maintaining competitiveness with private vehicles. The value of γ is chosen to prioritise serving feasible requests whenever marginal profit is non-negative, without forcing low-quality assignments. Requests are grouped every $\Delta t = 30s$ [19], and each request considers at most 20 candidate vehicles selected from nearby garages based on real-time travel times. With this pruning strategy, the resulting ILPs are solved to optimality within each batching interval, with an average computational time of less than 1 second. Note that the proposed ILP has the same totally unimodular structure as the classical linear assignment problem [19], which allows it to be relaxed to a linear program while still yielding integer optimal solutions. Consequently, the optimality gap is zero and the optimisation is solved efficiently within 1 s per batch; see Appendix B for detailed computational results.

The parameters in (8), (9), and (10) are set as $f^0_{pv} = 6$ €, $f^v_{pv} = 0.9\,\mathrm{km}/€$, $\beta_{pv} = \beta_{SAV} = 0$, $\beta_t = 0.48$, and $\beta_f = 3.2$, based on [5] and [23]. We adjusted the utility parameter for private vehicles, shifting it from $\beta_{pv} = -0.1$ for human-driven vehicles to $\beta_{pv} = 0$ for AVs. Furthermore, from (1), we employ $f_0 = 2.5$ €, $f_t = 0.2$ €/min ,$f_d = 0.6$ €/km [40].

All above parameter values are selected based on the relevant literature and adapted to reflect real-world operating conditions in Tallinn. In addition, sensitivity analyses are conducted for the key parameters to assess the robustness of the results.

4.2 The behavioral models

The initial pool of eligible requests, i.e., users willing to adopt SAV as a travel mode, is defined through activity-based behavioural modelling. A model of the population for the city of Tallinn is built in SimMobility MT [41], which takes as input a synthetic population of 400,000 individuals, then exploits nested logit and utility functions to compute each mobility choice within the simulated day through 22 behavioural models. Each individual is characterised by socio-demographic features such as income level, household structure, and number of cars owned, which are considered in the utility formulas through weights, denoted as β. More details about the employed activity-based model, its structure and calibration may be found in [42]. While, in this study, we focus on the supply side of the problem (i.e. the effects of different routing algorithms on the service performance) and thus make fixed assumptions on the demand, we refer the reader to [43] for additional details on the feedback effects of different service performances on the demand. In the following, we summarise how the utility of the SAV mode is framed. Before the start of the simulation, the reduction in price due to sharing is not considered, as the population is presented the SAV alternative through the following utility function:

$$\begin{aligned} U_{\mathrm{SAV}} = & \beta_{\mathrm{SAV}} + \beta_{\mathrm{cost}} \cdot \alpha \cdot \frac{cost_{\mathrm{SAV}}}{income} + \\ & \beta_{\mathrm{travel\_time}} \cdot travel\_time + \\ & \beta_{\mathrm{downtown}} \cdot dummy_{\mathrm{downtown}} + \\ & \beta_{\mathrm{population}} \cdot \log(\exp(\beta_{\mathrm{area}} \cdot area) + \\ & \exp(\beta_{\mathrm{density}} \cdot population)) + \beta_{\mathrm{distance}} \cdot distance, \end{aligned} \tag{11}$$

where β are the weights for each component of the utility function; area, density and population are instead aggregate measures for each OD pair, which represent the increased utility related to zones with higher population density (e.g., downtown). Eq. (11) represents the SAV service utilisation rate estimation and allows us to define the initial pool of potential users by comparing the SAV utility with the one of other modes of transport. In this study, the fare per kilometre of SAVs, i.e., $cost_{\mathrm{SAV}}$ in (11) is assumed to be half of the one currently available for human-driven taxis in Tallinn to account for the lack of operational cost related to the driver, similarly as in [33, 44]. Note that the effects of ridesharing, its performance (expressed through travel and waiting time), and the reduction in pricing are considered in a successive phase, i.e., when the vehicles are assigned to each request, which allows to correctly represent the detour and waiting times.

The activity-based model has been used to design a future scenario relevant to the research questions and resulted in a modal share of 21% for the SAV service, against 31% for private vehicles, 24% for public transport,

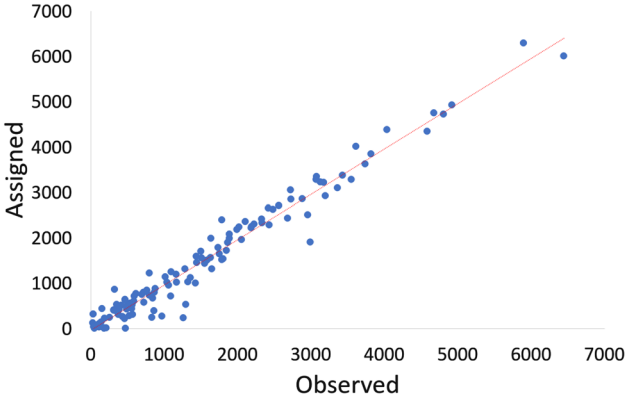


Fig. 3 Scatter plot with the baseline and simulated values of flows across the available detectors

23% for walking, and 1% for other modes (e.g., motorcycle, cycling). Public transport is not explicitly modelled as an operational alternative during the simulation. Instead, it is accounted for upstream through the activity-based demand model, where travellers choose among public transport, private vehicles, walking, and SAVs based on calibrated utilities. Only travellers selecting SAVs are passed to the simulation framework, while public transport demand remains fixed and contributes implicitly to background traffic conditions.

4.3 Morning peak scenario

4.3.1 Demand

The results from the activity-based model have then been analysed for the morning peak hours, i.e. 08:00–10:00, the period of the day when congestion is higher. The background traffic equals 47,000 private vehicles within the network, plus 22,500 commuter vehicles (still private) travelling from outside the city boundaries. There are 34,750 trip requests for SAV service and 47,000 private trips during the two-hour morning peak. Note that the travel demand generated by SimMobility lacks parameters, such as maximum waiting time and latest arrival time, thus we synthesise these data based on [23] and [8]. To capture heterogeneity in travellers' willingness to share, we assign request-specific maximum waiting and detour tolerances. For each request, the maximum waiting time is drawn from a truncated normal distribution with a mean of 7 minutes and bounded between 5 and 15 minutes, while the maximum detour tolerance is drawn from a truncated normal distribution with a mean of 10 minutes and bounded between 0 and 20 minutes. These tolerances are enforced as individual acceptance constraints during the matching process. Lastly, for requests that cannot be served by the SAV system, either due to violating the waiting time threshold or resulting in an unfair trip fare, we assume that a private vehicle performs a solo trip.

4.3.2 Supply

At the beginning of the simulation, 3,050 four-seat SAVs are uniformly distributed across the network, with five vehicles at each centroid. The SAV fleet movement is simulated in Aimsun Next, where it interacts with other human-driven vehicles. To achieve this, a dynamic traffic assignment model (DTA) has been built and calibrated for the city of Tallinn. The calibration has been carried out against data from 148 detectors (Fig. 3) and for the morning peak hours (07:00–10:00), when the network is the most congested. It focuses on the parameters of the volume-delay functions adopted by the simulator and achieves a R^2 value of 0.9625 and GEH < 5 for 75% of detectors [45], threshold on par with current best practices [46]. The routing relies on historic routing, which is calculated via a macroscopic static traffic assignment algorithm, for 70% of the background traffic, and on stochastic route choice for the remaining 30%.

4.3.3 Compared strategies

To evaluate how fairness-aware price-based optimisation affects the performance of DRS services and its influence on congestion, we designed five scenarios, which are summarised in Table 1, and more thoroughly described as follows.

- NRTX serves as the baseline scenario without ridesharing, where the nearest (NR) idle vehicle is assigned to serve each request, and the trip price is calculated via (1), representing a conventional taxi (TX) fare.
- TTDE indicates assigning the vehicle with the minimum travel time (TT) and utilising a detour-

Table 1 Compared assignment and pricing strategies

Strategy	Ridesharing	Assignment method	Pricing method
NRTX	No	Nearest idle vehicle (NR)	Conventional taxi fare (TX), computed by (1)
TTDE	Yes	Travel-time-based (TT) heuristic assignment	Detour-based pricing (DE), adapted from the mt-Share method [24]
TTFA	Yes	Travel-time-based (TT) heuristic assignment	Fairness-aware pricing (FA)
OPDE	Yes	Optimisation-based (OP) assignment via the profit-oriented optimisation problem (4)–(7)	Detour-based pricing (DE)
OPFA	Yes	Optimisation-based (OP) assignment	Fairness-aware pricing (FA)

based (DE) pricing method. TTDE enhances NRTX by allowing ridesharing, while the vehicle-traveller pairs are formed using a simple heuristic method, consisting of assigning the vehicle that results in the minimum total travel time after inserting the new traveller's origin and destination. Besides, the detour-based method is a common pricing method for dynamic ridesharing. Here we employ a modification of the mt-Share method proposed in [24]. However, instead of employing the entire framework for dynamic ridesharing proposed in [24], we utilise only the detour-based pricing method to fairly split the ridesharing benefit between co-riders.

- TTFA means travel time (TT) based assignment and fairness-aware (FA) pricing method. TTFA differs from TTDE in terms of the pricing method, where the former employs the same assignment strategy as the latter, but combined with the fairness-aware pricing method.
- OPDE represents optimisation-based (OP) assignment and detour-based (DE) pricing. OPDE extends TTDE, improving operation efficiency by incorporating the profit-oriented optimisation problem (4)-(7).
- OPFA indicates optimisation-based (OP) assignment and fairness-aware (FA) pricing. OPFA combines the optimisation assignment used in OPDE with the fairness pricing employed in TTFA.

These acronyms represent combinations of **matching** and **pricing** strategies. Specifically, the first two letters denote the matching method, e.g., whether it is heuristic matching (NR or TT) or optimisation-based (OP) matching; while the last two letters specify the pricing scheme, i.e., detour-based (DE) or fairness-aware (FA) pricing.

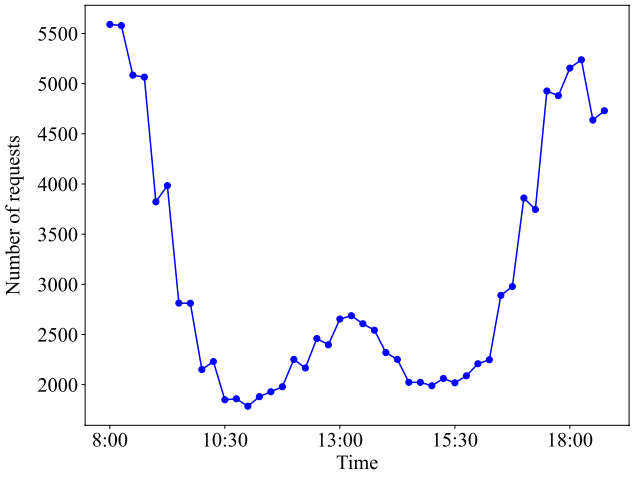


Fig. 4 Temporal distributions of trip requests in the full business day scenario

Finally, note that all scenarios utilise the same algorithm proposed in [6] to dispatch idle vehicles.

On the other hand, we compare OPFA with the two-fleet baseline scenario [8], where both ridehailing and ridesharing fleets are operated simultaneously, with 1,525 vehicles in each fleet. In contrast, this study relaxes the assumption of two distinct fleets and instead applies the same pricing method using a single-fleet-based matching approach for fleet operations. Besides, the simulations carried out in this study do not rely on historical travel times and instead allow the proposed algorithm to respond to dynamic changes in congestion patterns and travel times.

4.4 Full business day scenario

To comprehensively assess the ridesharing service over an extended period, we simulate the fleet operation from 8:00 to 19:00, covering both the morning and evening peaks as well as the non-peak noon hours. Similar to the peak scenario, the initial demand for this extended period is generated using the described activity-based model. The initial demand consists of 132,621 requests for SAVs, while the background traffic comprises 157,000 private vehicle trips. The temporal distribution of trip requests over the full business day scenario is displayed in Fig. 4. Instead of initialising vehicles homogenously across the network, we distribute the vehicles according to the spatial distribution of requests over the full day. In total, there are 4457 vehicles. In this scenario, we only test the best performance method OPFA to replicate the daily operation of DRS service providers. In this scenario, we set the discount parameter α to 0.4 to encourage ridesharing. Meanwhile, for the δ in Eq. (3) two values are tested (0.5 and 1), to understand how solo trip tax affect the use of ridesharing service.

5 Results and discussion

We first evaluate five single-fleet scenarios across three dimensions: fleet performance, service performance, and network congestion. Next, we examine the robustness of the results through a sensitivity analysis on key parameters. Finally, we compare the two-fleet baseline scenario from [8] with the best-performing OPFA strategy.

In terms of fleet performance, we measure the following metrics. The service rate is defined as the percentage of served requests. Total vehicle kilometres travelled (VKT) represents the total distance travelled by the entire fleet. Empty VKT refers to kilometres travelled without passengers on board and consists of two components. Pickup VKT measures the distance travelled by a vehicle from its current location to a passenger's pickup point after assignment, while Repositioning VKT captures the distance travelled by idle vehicles during repositioning actions when no passenger is on board. The key

Table 2 Fleet performance in the morning peak scenario

Strategy	Served requests	Total SAV	VKT/served	Empty VKT [km]		Average occupancy
		VKT [km]		Pickup	Repositioning	
NRTX	15274	239122	15.81	71994	8144	1
TTDE	13753(15729*)	228268	16.73	56716	12203	1.26
TTFA	15325	201513	13.72	55770	12394	1.38
OPDE	18041(6503*)	214430	12.18	60707	6582	1.49
OPFA	22789	206077	9.04	39494	6874	1.91

Note: *indicates the number of unfair pricing

Table 3 Service performance in the morning peak scenario

Strategy	Waiting time [min]			In vehicle time [min]		
	Mean	Median	stdev	Mean	Median	stdev
NRTX	6.07	4.69	6.42	16.53	12.5	13.99
TTDE	6.0	3.94	7.66	22.96	18.33	17.49
TTFA	5.24	3.70	6.48	23.11	17.67	18.97
OPDE	7.67	5.94	6.75	22.65	17.83	17.79
OPFA	6.20	4.72	6.55	24.09	20.17	18.0

distinction between the two is that Pickup VKT is associated with assigned requests, whereas Repositioning VKT arises from unassigned requests and reactive fleet rebalancing. Average occupancy is the average number of passengers on board during the simulation period. Besides, we calculate the statistics of passengers' waiting and in-vehicle time. Lastly, the network congestion is measured by average vehicle speed and density.

5.1 Morning peak scenario

5.1.1 Fleet performance

Table 2 reports the main indicator concerning the performance of the SAVs across the network and the served requests.

Table 2 allows us to isolate the individual effects of ridesharing, matching, and pricing. Comparing NRTX and TTDE identifies the pure effect of introducing ridesharing under heuristic matching and detour-based pricing: occupancy increases from 1 to 1.26, while empty VKT declines, but the number of served requests decreases due to limited matching efficiency. Holding matching fixed, the comparison between TTDE and TTFA isolates the pricing effect: fairness-aware pricing substantially increases the number of served requests (+11.4%) and reduces VKT per served request, indicating that pricing primarily affects user acceptance and pooling participation rather than vehicle utilisation. Holding pricing fixed, the comparison between TTDE and OPDE isolates the effect of optimisation-based matching, which yields a large increase in served demand (+31.2%), higher occupancy, and a sharp reduction in repositioning-related empty VKT, highlighting matching as the dominant operational driver. Finally, OPFA combines both effects, achieving the highest occupancy (1.91), the lowest empty VKT, and the largest service rate, confirming that pricing and matching are complementary and jointly necessary to fully realise the benefits of ridesharing.

The numbers in parentheses in the "served requests" column denote requests rejected due to unfair pricing, which occur only under detour-based pricing. As shown, the combination of heuristic matching and detour-based pricing results in nearly 45% of total travellers experiencing unfair pricing. In contrast to TTDE, the optimisation-based approach OPDE reduces the share of travellers experiencing unfair pricing to approximately 19% of total travellers; this same group subsequently leaves the system, and their exit is the primary driver of the performance gap between OPDE and OPFA.

In summary, OPFA delivers the strongest overall performance in Table 2, achieving the highest occupancy and the lowest empty VKT and VKT per served request. The occupancy gap relative to the second-best strategy (OPDE) is substantial (0.42), which directly translates into fewer empty pickup kilometres and improved traffic efficiency. OPDE ranks second across these indicators, confirming that optimisation-based matching is the dominant contributor to performance gains. Fairness-aware pricing alone (TTFA) also improves efficiency relative to TTDE, reducing VKT per served request and increasing served demand, indicating that pricing primarily enhances user acceptance and pooling participation.

5.1.2 Service performance

Table 3 reports indicators related to the SAV service, representing how effectively the SAV fleet performs while guided by the different algorithms. As the waiting and travel time increase for SAV, users become more likely to choose private vehicles. Differently from the initial modal choice model in SimMobility (described in 4.2), carried out with only an estimate of travel and waiting time at

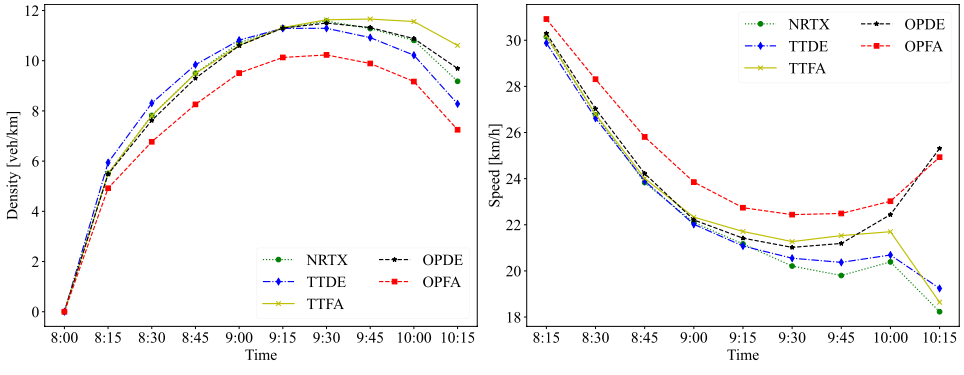
the start of the day, this choice happens during the traffic assignment and will impact the availability of the SAV vehicles in the following time steps. From Table 2, we observe that the number of travellers served by ridesharing increases through each scenario from NRTX to OPFA. As a result, all the statistics related to travellers' in-vehicle time increase accordingly with the increase of served requests. Additionally, OPDE, where the assignment of vehicles to requests follows the optimisation algorithm but the pricing is not fairly shared across the ridesharing users, exhibits the longest waiting time. This can be explained by the fact that the detour-based pricing method (TTFA) fails to generate proper ridesharing prices and, due to unfair pricing, travellers leave the system and opt for private cars instead. Consequently, OPDE has fewer served travellers. On the other hand, the detour-based pricing method might provide fewer ridesharing benefits, making it less attractive to travellers. Those who choose PVs based on their utility might remain in the waiting pool until the next round of optimisation. Consequently, OPDE results in the longest waiting time.

Another important metric for service performance is the experienced in-vehicle time. From NRTX to OPFA, the number of travellers using ridesharing increases due to optimisation and fairness pricing. While OPFA shows the highest travel time value (see Table 3), it should be noted how this is attributed to a higher percentage of shared trips, which inevitably cause a higher travel time due to detours (e.g., it can be noticed how NRTX, with no shared trips, has a much lower value of average travel time when compared to the other scenarios). Under TTFA, fairness-aware pricing makes shared rides more attractive, leading to higher acceptance rates and faster assignment, which reduces waiting times. However, this also results in a higher proportion of shared trips with larger detours, thereby increasing average in-vehicle travel time compared to TTDE. Nonetheless, when comparing OPFA with the ridesharing baseline TTDE, we observe that 65.7% more travellers are served, with, on average, only 0.2 minutes longer waiting time and 1.1 minutes longer in-vehicle time. This demonstrates that advanced DRS operation strategies, such as optimisation-based matching and pricing, can fundamentally improve system performance. When considered separately from each other (TTFA and OPDE scenarios), it is possible to notice how the fairness-aware pricing (TTFA) results in an increase of shared trips that is lower than the optimisation in OPDE and yet, the average travel time is higher than in OPDE. This would suggest that, while both scenarios overperform when compared to NRTX and TTDE, the improvements due to the optimisation algorithm are higher, both in fleet efficiency and in the total number of served requests.

5.1.3 Network congestion

We now proceed with analysing how the service affects the rest of the traffic and the overall congestion. Figure 5 reports how the average vehicle density and speed across the network progresses through the simulation. The average density is calculated at the end of each 30-minute time interval as the average value across each link in the network. It represents how many vehicles are present along the link per kilometre. Higher density represents higher occupancy of each link and thus is a proxy for the experienced congestion. Meanwhile, the average speed is calculated as the mean of the speeds of all moving vehicles in the same 30-minute time interval. A higher average speed indicates less congestion. All the progressions are the result of the same demand so the different trend for each scenario is solely the result of the service performance and the overall number of vehicles (both private and SAV) on the network.

Combining Fig. 5 (a) and (b), we observe that OPFA has consistently the lowest density, hence the average speed in OPFA outperforms that of all other scenarios. This indicates that efficient DRS operation strategies can greatly improve the traffic situation. Before 9:15, TTDE



(a) Average density (b) Average speed

Fig. 5 Trajectory of the average speed and density through the morning peak scenario

has the highest density and lowest speed due to unfair pricing, which incurs more private cars. However, as ridesharing trips gradually finish after 9:15, the density of TTDE decreases rapidly while the speed increases accordingly. Surprisingly, the density in TTFA surpasses the other scenarios in density after 9:15, which can be partly attributed to ridesharing trips requiring longer travel times, leading to more vehicles circulating in the latter half of the simulation. Notably, after 10:00, vehicle speeds increase under the optimisation-based strategies (OPDE and OPFA), while they decrease under the heuristic-based strategies (NRTX, TTDE, and TTFA). This phenomenon is attributed to the cool-down process during the last 15 minutes, with no new ride requests. In heuristic-based strategies, many of the previous waiting requests cannot be served due to high congestion and limited waiting time. Conversely, by leveraging optimisation methods, OPDE and OPFA prevent the loss of potential passengers and, consequently, fewer private cars are utilised. Additionally, as most SAVs reach their destinations and travel on secondary roads, vehicle speeds drop significantly. Meanwhile, the optimisation method ensures efficient operation even during the cool-down process in OPDE and OPFA. As vehicle density drops, vehicle speed increases.

Furthermore, separating assignment from pricing, as observed in TTFA, leads to higher density, validating the assertion that pricing under an assumed matching policy can result in inferior performance [47]. The higher density in TTFA and OPDE reveals that inappropriate DRS strategies might exacerbate traffic conditions. Nonetheless, OPFA consistently maintains lower density and higher speed throughout the simulation. This is probably the result of the highest number of shared trips across scenarios. This, coupled with the highest number of served requests (and consequently the lowest number of private vehicles), results in an overall reduction in vehicles deployed and travelling to serve the same demand. This result suggests that combining the proposed optimisation algorithm with fair pricing yields the biggest improvements, not only for the SAV fleet but also for the background traffic. Therefore, a joint optimised method is critical for the DRS operator to enhance both fleet performance and traffic efficiency.

Finally, the aggregate results are further assessed by analysing their spatial distribution. Figure 6 compares the densities for TTFA and OPFA (i.e., the best and worst one in terms of density) at 9:30, the time at which density reaches the highest value for most of the scenarios. As it can be noticed, while the critical points and bottleneck are similar, the propagation of queue fronts is way more severe in TTFA. The degradation of the traffic flow across the network seems also to impact more the background private traffic, as the SAV service retains similar travel time performance between TTFA and OPFA (see Table 3); yet, it should be remembered how TTFA sees more requests shifting to private vehicles, thus serving only those requests which maintain a competitive performance despite the surrounding traffic conditions. An analysis of the travel time distribution for background traffic reveals that, while the distribution itself is fairly similar, OPFA still over-performs as its average travel time for background traffic is the lowest among scenarios (Fig. 7).

It should be stressed how, in our experiments, the SAVs are as (in)efficient as human-driven vehicles in terms of driving behaviour. This means that the improvement in travel time for the background traffic is due solely to the more efficient routing and fewer VKT in OPFA, as every other boundary condition (SAV fleet, SAV potential demand, overall demand) is kept constant across scenarios.

5.1.4 Performance comparison

Finally, we compare the OPFA with the two-fleet scenario. As shown in Table 4, the single fleet achieved a significantly higher service rate, with 17.5% more requests served compared with the two-fleet scenario. This is because the doubled number of ridesharing vehicles increases passenger-carrying capacity. Despite a slight

(a) TTFA (b) OPFA

Fig. 6 Density values [veh/km] for each section in TTFA and OPFA at 9:30 in the morning peak scenario

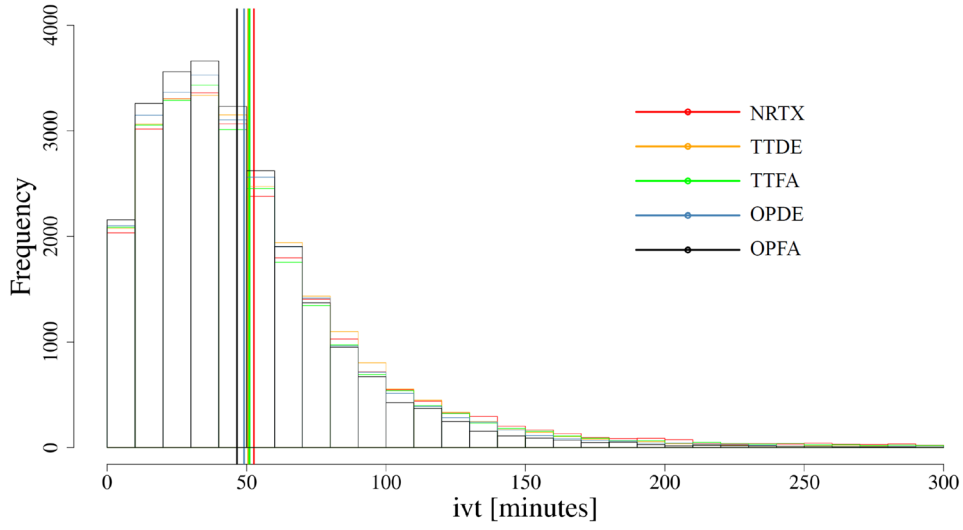


Fig. 7 Travel time distributions for the background private traffic in the morning peak scenario; averages are marked with the vertical lines

Table 4 Performance comparison in the morning peak scenario

Strategy	Served requests	VKT [km]		Average occupancy	Average time[min]	
		Total SAV	Empty		Waiting	Traveling
Two-fleet	16698	165201	77526	1.37	5.96	21.60
OPFA	22789	206077	46368	1.91	6.20	24.09

Table 5 Sensitivity of system performance to batching interval Δt (strategy OPFA)

Δt [s]	Served requests	Total VKT [km]	Empty VKT [km]		Service performance on average		
			Pickup	Repositioning	Waiting [min]	In-vehicle [min]	Occupancy
15	21,624	210,260	39,019	7205	7.0	25.3	2.07
30	22,789	206,077	39,624	7452	6.2	24.1	2.05
60	21,079	207,590	41,541	9645	6.6	24.9	2.03
120	20,006	210,913	42,638	12,407	6.6	25.4	1.92
300	17,182	206,896	50,742	11,886	8.1	23.9	1.62

increase in waiting and travel times, OPFA is able to serve significantly more passengers with only a moderate rise in VKT. Another observation is that overlooking dynamic travel times may lead to overly optimistic conclusions. Although we did not directly compare the effects of static and dynamic travel times on fleet operations using the same network, demand, and supply, the relative metrics in the New York City and Tallinn experiments, such as the ratio of requests to vehicles, suggest that dynamic travel times result in lower service rates. This finding underscores the importance of employing advanced traffic simulators to evaluate state-of-the-art methods for on-demand service operations.

5.2 Sensitivity analysis

5.2.1 Batching time

Table 5 presents the sensitivity of system performance to the batching interval Δt under the OPFA strategy. Overall, the results indicate that system performance is maximised at intermediate batching intervals, while excessively large batching intervals lead to a gradual deterioration in both service quality and fleet efficiency. For batching intervals up to 60 s, variations across most performance indicators remain relatively small, suggesting that OPFA is robust to moderate changes in Δt. As the batching interval increases, the average vehicle occupancy decreases, indicating that shorter batching intervals facilitate more effective ridesharing by enabling better consolidation of requests across successive batches. This higher level of sharing is accompanied by slightly higher waiting and in-vehicle times, reflecting the additional detours inherent to shared rides. When Δt exceeds 60 s, the number of served requests declines progressively, while waiting times increase and occupancy decreases, showing a degradation in service performance. Meanwhile, total VKT remains relatively stable for Δt values between 15 s and 120 s; however, empty VKT, particularly associated with vehicle repositioning, rises substantially at larger batching intervals, pointing to less effective fleet utilisation. In short, these results highlight

Table 6 Sensitivity of system performance to fleet size

Strategy	Fleet Size	Served Requests	Total VKT [km]	Empty VKT [km]		Service performance on average		
				Pickup	Repositioning	Waiting [min]	In-vehicle [min]	Occupancy
TTDE	1220	7568	210,179	22,494	3461	8.83	25.2	1.56
	3050	13,753	228,268	56,716	12,203	6.0	23.0	1.26
	5490	15,994	221,987	73,895	26,441	7.9	24.0	1.23
OPFA	1220	8776	202,235	16,651	2237	8.72	26.6	2.03
	3050	22,789	206,077	39,494	6874	6.2	24.1	1.91
	5490	31,369	227,902	58,751	23,664	5.47	23.4	1.96

Table 7 Fleet performance in the full business day scenario

Penalty δ	Service rate [%]	Total SAV VKT [km]	VKT per request	Empty VKT [km]		Average		
				Pickup	Repositioning	Occupancy	Time idle [mins]	Fare [€]
1	95.5	785650	6.2	386393	4400	1.36	104	6.8
0.5	94.0	685139	5.5	215988	9856	1.55	137	5.9

a clear trade-off between responsiveness and coordination efficiency, with $\Delta t = 30$ s providing the best balance between service performance and operational efficiency under OPFA.

5.2.2 Fleet size

Table 6 reports a sensitivity analysis of system performance with respect to fleet size for the two representative strategies, TTDE and OPFA. Three fleet sizes are considered, ranging from an undersupplied case to an oversized fleet, while keeping demand unchanged. Several consistent patterns emerge. First, increasing fleet size naturally leads to a higher number of served requests and lower waiting times for both strategies; however, the relative performance ranking remains unchanged across all fleet sizes. The roughly two-minute higher waiting time for TTDE with 5490 vehicles reflects lower passenger acceptance, triggering additional empty vehicle repositioning and, consequently, inefficient fleet utilisation. In particular, OPFA consistently serves substantially more requests than TTDE while maintaining comparable or lower total VKT, confirming that the optimisation-based matching combined with fairness-aware pricing scales robustly with fleet capacity. Second, while empty VKT increases with fleet size for both strategies, reflecting higher repositioning and pickup activity in looser supply conditions, OPFA systematically meaningfully lower empty VKT per served request, indicating more efficient vehicle utilisation. Third, average occupancy declines slightly as fleet size increases, yet OPFA preserves a significantly higher occupancy level across all scenarios, suggesting that its ability to consolidate demand into shared trips is not sensitive to fleet abundance. Overall, these results demonstrate that the advantages of OPFA are not driven by a specific fleet-size calibration, but persist across a wide range of supply conditions, thereby reinforcing the robustness and external validity of the comparative findings.

5.3 Full business day scenario

Since OPFA outperforms the other algorithms tested in the morning peak scenario, it is worth investigating how it performs throughout an entire business day of on-demand system operation. As the objective of the optimisation problem is to maximise trip profit, we analyse how penalising solo trip profits affects system performance. Particularly, we test two scenarios, where in the baseline scenario S1 we set $\delta = 1$ in Eq. 3, i.e., no tax is imposed on solo trips; this is compared with a taxed scenario S2, where we set $\delta = 0.5$ in Eq. 3 so as to prioritise ridesharing trips.

Table 7 compares the fleet performance of OPFA for scenarios S1 and S2 over the 11-hour period. Despite 1% difference in service rate, a smaller δ results in significantly less VKT, approximately 12.8% less, due to increased ridesharing. Hence, the average VKT per served request is lower in S2. Interestingly, we found that the reduction in VKT primarily comes from the empty VKT associated with travelling to the requests' origins. Besides, since S2 imposes a higher solo trip tax, the operator prioritises ridesharing trips. Consequently, S2 shows a higher occupancy rate and idle times but lower average trip fares.

Figure 8 presents the average number of passengers on board across the simulation period. As shown in Fig. 4, the number of requests sharply decreases from 8 to 10 AM. Correspondingly, the average number of passengers on board first increases sharply and then decreases rapidly. During the non-peak hours, the average number of passengers in S1 and S2 is 1 and 1.4, respectively. Both begin to rise again during the evening peak. It is worth noting that both curves have the same number of passengers during the peak hours, as there are no enough idle vehicles available. When the fleet operates at full capacity, the impact of different choices of δ becomes insignificant. The main difference lies in the off-peak hours, where a smaller δ allows the operator to assign more

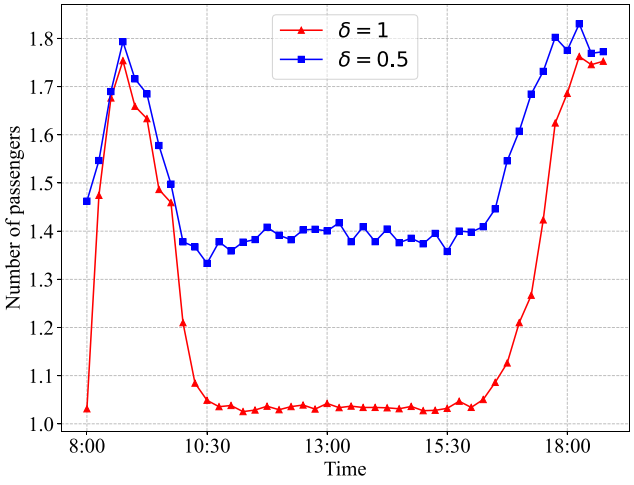


Fig. 8 Average number of passengers in vehicles in the full business day scenario

passengers to en-route vehicles instead of idle ones. This greatly reduces the empty driving VKT and increases the occupancy rate.

Figure 9 shows the average waiting and travel times throughout the simulation period. Both metrics exhibit similar trends: they increase during the morning peak, drop during non-peak hours, and rise again during the evening peak. However, the travel time decreases sharply at the end of the simulation. This is because long trips or newly assigned trips at the end of the simulation are not included in the analysis (i.e. trips generated during the last time interval, with not enough time to carry out their natural progression before the end of the simulation). Additionally, when $\delta = 1$, passengers experience longer waiting times but shorter in-vehicle times. This occurs because passengers wait longer for available ridehailing vehicles, while ridesharing offers more ride options. As a result, with more ridesharing trips, waiting times decrease but in-vehicle travel times increase. This further suggests that the S2 scenario has a higher proportion of ridesharing trips. However, the total travel times in both scenarios are quite similar.

5.4 Discussion

The operational efficiency of different scenarios has several policy implications involving both MoD service operators and public authorities. For the operator, DRS is capable of serving significantly more travellers, even when considering real-time traffic conditions. However, this improvement is achieved through advanced operational strategies. Optimisation-based assignment could significantly increase the number of served travellers, whereas simple heuristic methods, such as assigning the vehicle with the minimum travel time, may even reduce the number of served travellers. Besides fair pricing could further incentivize ridesharing users, thereby improving overall performance. Lastly, with the same fleet size, advanced DRS operational strategies could serve more

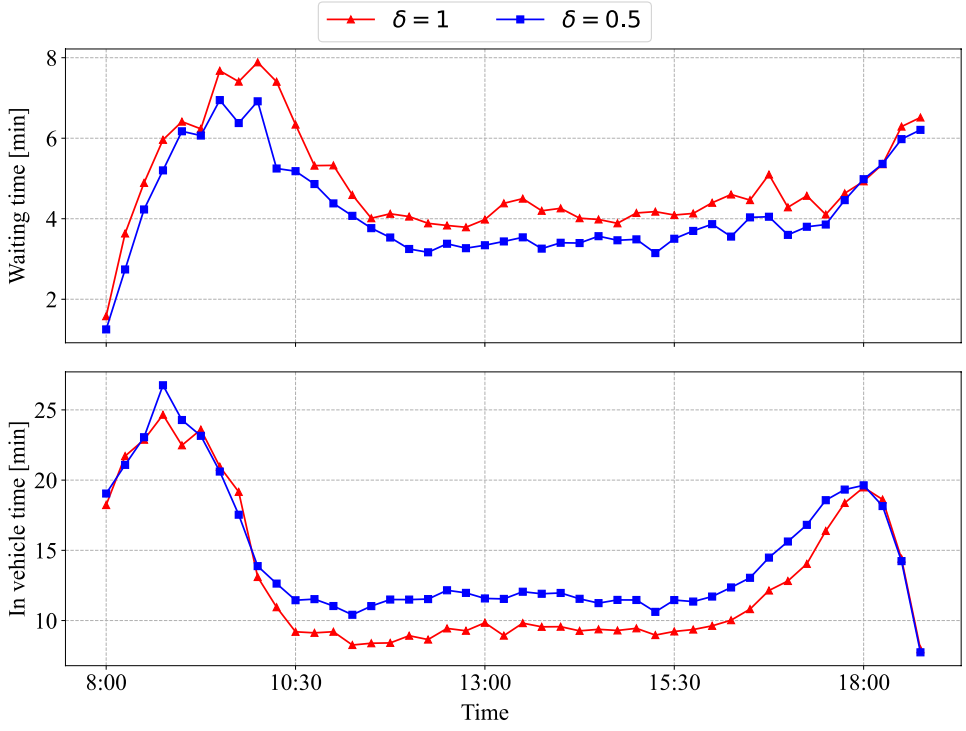


Fig. 9 Passenger's average waiting and travel time across the full business day scenario

travellers with even less VKT, thereby reducing operational costs and increasing profits. Therefore, the service operator must deploy advanced DRS operation strategies to improve system performance while maintaining good service quality. This requires designing algorithms also from the traveller's perspective. Conventional private operators typically prioritise self-interest and profit. Consequently, they may overlook the congestion effects brought about by their fleets. However, our research finds that a strategy solely focused on profit, without integrating matching and pricing, can lead to worse outcomes. Being aware of the potential congestion effects caused by the fleet and taking proactive measures to optimise its operation not only benefits other modes of travel but also proves advantageous for the operator itself.

On the other hand, private trips, no matter whether implemented by PVs or SAVs, can lead to severe congestion and emission problems. The public sector can encourage travellers to use ridesharing services by introducing higher fees on private trips or imposing lower taxes on DRS services. Given the potential for reducing congestion and emissions through DRS, the public sector could act as the service operator to encourage the adoption of this more sustainable mode of travel. Additionally, public-owned MoD services, prioritising societal benefits over profit, have the potential to further optimise DRS operations with reduced trip fares.

In the full business day scenario, we observed that a higher solo trip tax has a positive effect on promoting ridesharing services. As solo trips, such as conventional ride-hailing, are generally more profitable for the service operator. This implies that if policymakers leave service providers unintervened, their preference for solo trips would lead to significantly higher VKT and thus severe congestion and emission problems. Transport authorities should raise concerns about the negative effects brought by the profit-oriented on-demand service. Moreover, they should erect policies, such as imposing additional tax on solo trips to promote the use of ridesharing modes. As revealed in the full business day scenario, ridesharing can significantly reduce VKT and thus the associated external effect.

In our framework, the fairness-aware pricing mechanism explicitly ensures that no traveller pays more for a shared trip than for the corresponding solo trip, and that users experiencing larger detours or longer shared distances receive greater fare reductions. This design particularly benefits more price-sensitive users, including lower-income travellers, and encourages participation in ridesharing. From an accessibility perspective, the optimisation-based matching and pricing strategy increases overall service coverage and the number of served requests, which can improve access to mobility for users in lower-demand or peripheral areas where conventional taxi is less reliable. While we do not explicitly segment users by income in the real-time pricing stage, the demand generation relies on an activity-based model calibrated with socio-demographic attributes, meaning that lower-income users are more likely to enter the system when prices are competitive.

6 Conclusions

This study evaluates the impact of an advanced DRS pricing and matching method on urban traffic using a mesoscopic traffic simulator while explicitly modelling users' mode choices. Building on [8], we relax the key assumption about historic travel times and test the proposed algorithm with dynamic travel times. We see that dynamic travel times present additional challenges to on-demand service operations, resulting in less favorable performance compared to the static case of [8]. This underscores the importance of testing new methods with advanced traffic simulators and accounting for traffic dynamics. The study also reports valuable results about the impacts of the developed methods for a medium sized European capital. SimMobility is utilised to synthesise the future demand for SAV. We then compare the optimal assignment and fairness pricing with other heuristic methods and test various scenarios in a real urban network. Compared to heuristic methods (TTDE), the optimisation-based matching (OPDE) can serve 31.2% more travellers with 6.1% less VKT and 1.67 minutes longer waiting time on average. Furthermore, comparing OPDE with OPFA, 26.3% more travellers are served with 3.9% less VKT and 1.4 minutes less waiting time. This demonstrates that accounting for travellers' mode choices and pricing fairness could further enhance the performance of such a DRS system. On the other hand, OPFA has been proven to significantly reduce vehicle density and thus lead to higher travel speeds and less congestion. The full business day scenario shows the robustness of the OPFA method, indicating its applicability in real-time city-scale ridesharing. The simulation results reveal that advanced operation strategies are critical for successful DRS implementation. On the other hand, rule-based methods are unable to exploit the potential of DRS and thus limit the utilisation of DRS. Contrary to intuition, ridesharing (TTDE and TTFA) does not always lead to better transport efficiency. This work demonstrates that ridesharing services need to be better organised to improve both fleet performance and reduce network congestion.

Furthermore, even if the same strategy is implemented, different choices of certain key parameters can lead to dramatically different outcomes, as demonstrated in our full business day scenario. For profit-oriented service operators, transport authorities should promote the use of a more sustainable mode - ridesharing, by either taxing solo trips or subsidising ridesharing trips.

Finally, it is important to acknowledge the limitations of this study. First, the numerical results are specific to the analysed city and fleet sizes; thus, the findings primarily provide comparative and mechanism-based insights, rather than externally valid performance benchmarks across cities or fleet scales. In addition, while we model private vehicles and SAVs, public transport was not explicitly included as a dynamic mode alternative in the traffic assignment mode choice. Future extensions should integrate public transit to assess how SAVs complement or compete with mass transit in real-time. Moreover, the study operates on a fixed demand generated by the activity-based model; it does not capture the long-term feedback loop where improved SAV service might fundamentally alter vehicle ownership rates or induce new demand. Besides, the adoption of SAVs depends on behavioural factors such as social acceptance and trust. While our models account for heterogeneous traveller preferences, the results should be interpreted as conditional on these behavioural assumptions and the chosen fare levels. Future research may also compare alternative policy instruments, such as distance-based solo surcharges or occupancy-based discounts. While the tax-based formulation provides a transparent benchmark, broader policy comparisons are left for future work. Lastly, we simulated a single-operator monopoly, whereas a real-world scenario might involve competition between multiple service providers. Given the Aimsun Ride allows for multiple operators, it would be interesting to investigate the competition and cooperation strategies between similar service providers.

Appendix A Fairness pricing discount function for ridesharing trips

Following [8], we define two parameters to reflect ridesharing trip features, including detour distance, the number of co-rides, and shared trip distance. The detour index τ_r^{d} is calculated via the following equation:

$$\tau_r^{\mathrm{d}} = \frac{d_r^{\mathrm{u}} - d_r^{\mathrm{s}}}{d_r^{\mathrm{u}}}, \tag{A1}$$

where d_r^{u} indicates the updated trip distance after route insertion, while d_r^{s} denotes the vehicle path length before insertion. On the other hand, we define ζ_r^{g} as the ridesharing index which reveals the distance and the number of passengers of shared rides experienced by traveller r. To calculate ζ_r^{g}, the vehicle path is divided into legs by passenger's pick-up and drop-off locations. For each leg λ, we first calculate the travel fare by $f_\lambda^{\mathrm{h}} = f_t \cdot t_\lambda + f_d \cdot d_\lambda$, where f_t, t_λ, f_d, and d_λ denote unit time fee, leg travel time, unit distance fee, and leg distance, respectively. Nonetheless, if a leg is shared by several passengers, the leg fare f_λ is divided by the number of passengers k_λ. Hence the *leg shared fare* $f_\lambda^{\mathrm{s}} = f_\lambda^{\mathrm{h}}/k_\lambda$. Next, a binary parameter $y_{p,\lambda}$ is introduced to identify if passenger p travels through leg λ. Lastly, we define

$$\zeta_r^{\mathrm{g}} = 1 - \frac{\sum_\lambda f_\lambda^{\mathrm{s}} * y_{r,\lambda}}{\sum_\lambda f_\lambda^{\mathrm{h}} * y_{r,\lambda}} \tag{A2}$$

as the ridesharing index, which can be used to measure the degree of ridesharing, such as the number of coriders and the length of shared travelled distance. Finally, the ridesharing discount rate is defined as:

$$\Phi_r = \exp(-\alpha(\tau_r^{\mathrm{d}} + \zeta_r^{\mathrm{g}})) - b, \tag{A3}$$

where the discount coefficient α is designed to control the discount rate. Besides, a based rate b is defined so that $b > 0$ if the vehicle is empty and $b = 0$ otherwise. In our experiments, we set $\gamma = 2$ €/pax, $\delta = 0.7$, and discount parameter $\alpha = 0.2$ as in [8].

Appendix B Computational time

Figure B1 illustrates the per-batch computational time of the OPFA over the morning-peak simulation horizon, where each time step corresponds to one request batch of length $\Delta = 30s$. The solid blue line reports the total per-batch runtime, while the orange line indicates the time spent solving the batch-level assignment optimisation.

On average, the total runtime equals 9.08 s per batch, which remains well below the batch length throughout the simulation. The maximum observed runtime per batch is 29 s. The optimisation step itself requires only 0.87 s per batch on average, accounting for a negligible share of the overall computational effort. This demonstrates that the assignment ILP is solved efficiently to optimality and is not the computational bottleneck of the framework; most runtime is instead associated with candidate vehicle searching and route feasibility checks. We further note that the implementation is written in Python and has not been performance-optimised.

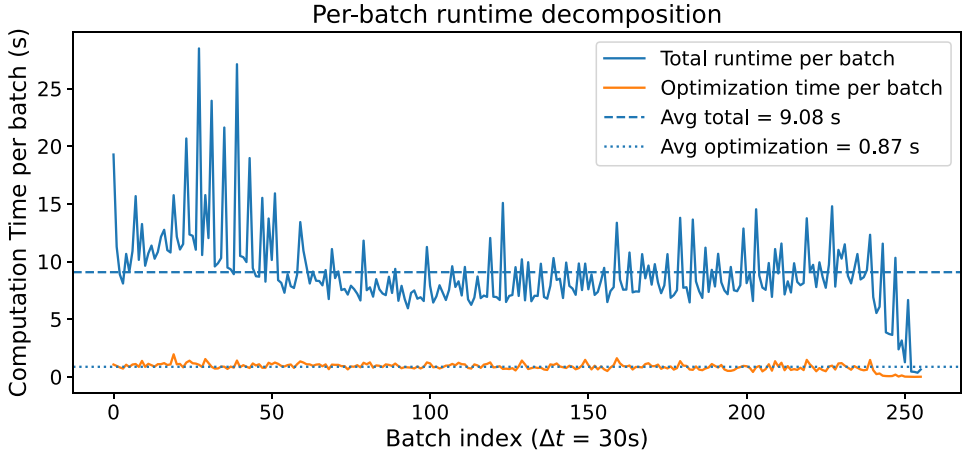


Fig. B1 Per-batch total computational time versus optimisation time

Acknowledgements
The work by the authors affiliated with Aalto University was funded by the Academy of Finland project ALCOSTO (no. 349,327).

Author contributions
Ze Zhou & Serio Agriesti: Conceptualisation, Methodology, Software, Validation, Formal analysis, Writing - original draft, Visualisation. Claudio Roncoli: Conceptualisation, Methodology, Writing - review & editing, Supervision, Funding acquisition. Lampros Yfantis & Jordi Casas: Methodology, Software, Writing - original draft; Bat-hen Nahmias-Biran: Conceptualisation, Writing - review & editing.

Funding
Open access funding provided by Chalmers University of Technology.

Data availability
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethical approval
The research meets all the applicable ethical standards concerning research integrity,moreover the paper is not result of duplicate, fraud or plagiarism.

Declaration
Our paper reports original research and is not submitted / published in any journal and is not being considered for publication elsewhere. All authors have seen and approved the manuscript and have contributed significantly for the paper.

Conflict of interest
None of the authors of this paper has a financial or personal relationship with other people or organizations that could inappropriately influence or bias the content of the paper.

Received: 12 December 2024 / Accepted: 24 March 2026
Published online: 03 April 2026

References

1. Luo, H., Bao, Z., Choudhury, F. M., & Culpepper, J. S. (2019). Dynamic ridesharing in peak travel periods. *IEEE Transactions on Knowledge and Data Engineering*, *33*(7), 2888–2902.
2. Olayode, I. O., Severino, A., Alex, F. J., Macioszek, E., & Tartibu, L. K. (2023). Systematic review on the evaluation of the effects of ride-hailing services on public road transportation. *Transportation Research Interdisciplinary Perspectives*, *22*, 100943.
3. Furuhata, M., Dessouky, M., Ordóñez, F., Brunet, M.-E., Wang, X., & Koenig, S. (2013). Ridesharing: The state-of-the-art and future directions. *Transportation Research Part B: Methodological*, *57*, 28–46.
4. Loeb, B., & Kockelman, K. M. (2019). Fleet performance and cost evaluation of a shared autonomous electric vehicle (SAEV) fleet: A case study for Austin, Texas. *Transportation Research Part A Policy and Practice*, *121*, 374–385.
5. Gurumurthy, K. M., Kockelman, K. M., & Simoni, M. D. (2019). Benefits and costs of ride-sharing in shared automated vehicles across Austin, Texas: Opportunities for congestion pricing. *Transportation Research Record*, *2673*(6), 548–556.
6. Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., & Rus, D. (2017). On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences*, *114*(3), 462–467.
7. Hyland, M., & Mahmassani, H. S. (2018). Dynamic autonomous vehicle fleet operations: Optimization-based strategies to assign AVs to immediate traveler demand requests. *Transportation Research Part C: Emerging Technologies*, *92*, 278–297.
8. Zhou, Z., Roncoli, C., & Sipetas, C. (2023). Optimal matching for coexisting ride-hailing and ridesharing services considering pricing fairness and user choices. *Transportation Research Part C: Emerging Technologies*, *156*, 104326.
9. Räth, Y. M., Balac, M., Hörl, S., & Axhausen, K. W. (2023). Assessing service characteristics of an automated transit on-demand service. *Journal of Urban Mobility*, *3*, 100038.
10. Levin, M. W., Kockelman, K. M., Boyles, S. D., & Li T. (2017). A general framework for modeling shared autonomous vehicles with dynamic network-loading and dynamic ride-sharing application. *Computers, Environment and Urban Systems*, *64*, 373–383.
11. Narayanan, S., Chaniotakis, E., & Antoniou, C. (2020). Shared autonomous vehicle services: A comprehensive review. *Transportation Research Part C*, *111*, 255–293.
12. Narayanan, S., Chaniotakis, E., & Antoniou, C. (2022). Modelling reservation-based shared autonomous vehicle services: A dynamic user equilibrium approach. *Transportation Research Part C: Emerging Technologies*, *140*.
13. Berbeglia, G., Cordeau, J.-F., & Laporte, G. (2010). Dynamic pickup and delivery problems. *European Journal of Operational Research*, *202*(1), 8–15.
14. Toth, P., & Vigo, D. (2014). *Vehicle routing: Problems, methods, and applications*. SIAM, USA.
15. Agatz, N., Erera, A. L., Savelsbergh, M. W., & Wang, X. (2011). Dynamic ride-sharing: A simulation study in metro Atlanta. *Procedia-Social and Behavioral Sciences*, *17*, 532–550.
16. Horn, M. E. (2002). Fleet scheduling and dispatching for demand-responsive passenger services. *Transportation Research Part C: Emerging Technologies*, *10*(1), 35–63.
17. Ma, S., Zheng, Y., & Wolfson, O. (2013). T-share: A large-scale dynamic taxi ride-sharing service. In *2013 IEEE 29th International Conference on Data Engineering (ICDE)* (pp. 410–421.
18. Cordeau, J.-F. (2006). A branch-and-cut algorithm for the dial-a-ride problem. *Operations Research*, *54*(3), 573–586.

19. Simonetto, A., Monteil, J., & Gambella, C. (2019). Real-time city-scale ridesharing via linear assignment problems. *Transportation Research Part C: Emerging Technologies*, *101*, 208–232.
20. Zhou, Z., & Roncoli, C. (2022). A scalable vehicle assignment and routing strategy for real-time on-demand ridesharing considering endogenous congestion. *Transportation Research Part C: Emerging Technologies*, *139*, 103658.
21. Foti, L., Lin, J., & Wolfson, O. (2021). Optimum versus nash-equilibrium in taxi ridesharing. *GeoInformatica*, *25*, 423–451.
22. Fielbaum, A., Kucharski, R., Cats, O., & Alonso-Mora, J. (2022). How to split the costs and charge the travellers sharing a ride? Aligning system's optimum with users' equilibrium. *European Journal of Operational Research*, *301*(3), 956–973.
23. Jiao, G., & Ramezani, M. (2022). Incentivizing shared rides in e-hailing markets: Dynamic discounting. *Transportation Research Part C: Emerging Technologies*, *144*, 103879.
24. Liu, Z., Gong, Z., Li, J., & Wu, K. (2021). mT-share: A mobility-aware dynamic taxi ridesharing system. *IEEE Internet of Things Journal*, *9*(1), 182–198.
25. Azadeh, S. S., Atasoy, B., Ben-Akiva, M. E., Bierlaire, M., & Maknoon, M. (2022). Choice-driven dial-a-ride problem for demand responsive mobility service. *Transportation Research Part B: Methodological*, *161*, 128–149.
26. Horni, A., Nagel, K., & Axhausen, K. (Eds.). (2016). *Multi-agent transport simulation MATSim* (p. 618). London: Ubiquity Press. https://doi.org/10.5334/baw.
27. Fagnant, D. J., & Kockelman, K. M. (2018). Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in austin, texas. *Transportation*, *45*, 143–158.
28. Lokhandwala, M., & Cai, H. (2018). Dynamic ride sharing using traditional taxis and shared autonomous taxis: A case study of NYC. *Transportation Research Part C: Emerging Technologies*, *97*, 45–60.
29. Vosooghi, R., Puchinger, J., Jankovic, M., & Vouillon, A. (2019). Shared autonomous vehicle simulation and service design. *Transportation Research Part C: Emerging Technologies*, *107*, 15–33.
30. Basu, R., Araldo, A., Akkinepally, A., Basak, K., Seshadri, R., Nahmias-Biran, B., Deshmukh, N., Kumar, N., Azevedo, C. L., & Ben-Akiva, M. (2018). Automated mobility-on-demand vs. Mass transit: A multi-modal activity-driven agent-based simulation approach. *Transportation Research Record*, *2672*(8), 608–618.
31. Oke, J., Aboutaleb, M. Y., Akkinepally, A., Azevedo, C. L., Han, Y., Zegras, P. C., Ferreira, J., & Ben-Akiva, M. (2019). A novel global urban typology framework for sustainable mobility futures. *Environmental Research Letters*, *14*(9).
32. Nahmias-Biran, B., Oke, J. B., Kumar, N., Azevedo, C. L., & Ben-Akiva, M. (2021). Evaluating the impacts of shared automated mobility on-demand services: An activity-based accessibility approach. *Transportation*, *48*, 1613–1638.
33. Nahmias-Biran, B.-H., Dadashev, G., & Levi, Y. (2023). Sustainable automated mobility-on-demand strategies in dense urban areas: A case study of the tel aviv metropolis in 2040. *Sustainability*, *15*(22).
34. Wolf, F., Engelhardt, R., Zhang, Y., Dandl, F., & Bogenberger, K. (2023). Effects of dynamic and stochastic travel times on the operation of mobility-on-demand services. arXiv preprint arXiv:2308.05535.
35. Aimsun. (2023). Aimsun Next 23 User's Manual, Aimsun next 23.0.0 edn. (2023). Barcelona, Spain. https://docs.aimsun.com/next/23.0.0/.
36. Casas, J., Ferrer, J. L., Garcia, D., Perarnau, J., & Torday, A. (2010). Traffic simulation with Aimsun. In *Fundamentals of traffic simulation*. Springer.
37. Gipps, P. G. (1981). A behavioural car-following model for computer simulation. *Transportation Research Part B: Methodological*, *15*(2), 105–111.
38. Gipps, P. G. (1986). A model for the structure of lane-changing decisions. *Transportation Research Part B: Methodological*, *20*(5), 403–414.
39. Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand* (Vol. 9). MIT press.
40. Amigo Taxi Design: Amigo Taxi Design. (2023). https://en.amigotakso.ee/.
41. Lu, Y., Basak, K., Carrion, L. H. C., Adnan, M., Pereira, F. C., Saber, V. H., & Ben-Akiva, M. (2015). SimMobility mid-term simulator: A state of the art integrated agent based demand and supply model. In *Transportation Research Board 94th Annual Meeting*.
42. Agriesti, S., Kuzmanovski, J., Hollmén, J., Roncoli, C., & Nahmias-Biran, B.-H. (2023). A bayesian optimization approach for calibrating large-scale activity-based transport models. *IEEE Open Journal of Intelligent Transportation Systems*.
43. Agriesti, S., Roncoli, C., & Nahmias-Biran, B. H. (2025). A simulation-based framework for quantifying potential demand loss due to operational constraints in automated mobility services. *Transportation Research Part A Policy and Practice*, *192*, 104372.
44. Nahmias-Biran, B.-H., Dadashev, G., & Levi, Y. (2022). Demand exploration of automated mobility on-demand services using an innovative simulation tool. *Ieee Open Journal of Intelligent Transportation Systems*, *3*, 580–591.
45. Agriesti, S., Anashin, P., Roncoli, C., & Nahmias-Biran, B.-H. (2023). Integrating activity-based and traffic assignment models: Methodology and case study application. In *8th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems* (pp. MT-ITS.
46. Beeston, L., Blewitt, R., Bulmer, S., & Wilson, J. (2021). Traffic modelling guidelines. https://content.tfl.gov.uk/traffic-modelling-guidelines.pdf.
47. Özkan, E. (2020). Joint pricing and matching in ride-sharing systems. *European Journal of Operational Research*, *287*(3), 1149–1160.

Publisher's Note