Transportation Research Part A 190 (2024) 104289

Contents lists available at ScienceDirect

Transportation Research Part A

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

Improving public transportation via line-based integration of on-demand ridepooling

Andres Fielbaum [a,*], Alejandro Tirachini [b,c], Javier Alonso-Mora [d]

[a] School of Civil Engineering, University of Sydney, Australia
[b] Department of Civil Engineering and Management, University of Twente, Enschede, The Netherlands
[c] Department of Civil Engineering, Universidad de Chile, Santiago, Chile
[d] Department of Cognitive Robotics, Delft University of Technology, The Netherlands

ARTICLE INFO

Keywords:
Ridepooling
Public transportation
Transit networks
Quality of service

ABSTRACT

Ride-sourcing companies have worsened congestion in numerous cities worldwide, as many users are attracted from more sustainable modes. To reverse this trend, it is crucial to leverage the technology of connecting users and vehicles online and use it to strengthen public transport, which can be achieved by integrating on-demand pooled services with existing fixed-line services. We propose an efficient and practical integration idea: namely, to complement fixed bus lines with a fleet of smaller vehicles that follow flexible (on-demand) routes side-by-side with the fixed routes, so that part of the demand that would have used the fixed line can ride the flexible service instead. With this scheme, a smaller bus fleet is required, partially compensating for the increase in operators' costs stemming from the flexible vehicles. This integration strategy favors mostly two types of users: those traveling in low-demand periods, through lower waiting times, and those located far from the bus stops, because the on-demand vehicles can reduce their access time. We develop simulations in real-world scenarios from Santiago, Chile, and Berlin, Germany, for the cases of human-driven and automated vehicles. Results show that when vehicles are automated: (i) A small number of on-demand vehicles can reduce average walking times from approximately 12 to 2 min while reducing operators' costs, leading to a Pareto improvement, (ii) A larger number of on-demand vehicles can diminish total costs by 13%–39%, through a reduction in users' costs, although increasing operators' costs. If vehicles are not automated, total costs are reduced by more than 10% in all of the scenarios analyzed, but a Pareto improvement is not always possible. In general, this mixed fixed/on-demand system outperforms the use of on-demand ridepooling only. Results are more promising in Berlin, because large buses are cheaper in Santiago and run more crowded, so it is more costly to partially replace them by smaller vehicles.

1. Introduction

1.1. Motivation

Several of the major challenges faced by cities around the globe relate to the increasing trends in private vehicles ownership and usage. Congestion, emissions, urban sprawl, allocation of public space for parking and roads, and unequal commuting times are all related to this same cause (Baum-Snow, 2007; Gössling et al., 2016; Kenworthy, 2003; Will et al., 2020; Lucas, 2012). The

* Corresponding author.
E-mail address: andres.fielbaum@sydney.edu.au (A. Fielbaum).

https://doi.org/10.1016/j.tra.2024.104289
Received 16 February 2023; Received in revised form 27 September 2024; Accepted 12 October 2024
Available online 1 November 2024

emergence of transportation network companies (TNCs) that provide so-called ride-sourcing or ride-hailing services, in which several passengers can board sequentially the same vehicle, once promised to help relieve these problems thanks to a more intensive use of vehicles. However, now several studies have shown that ride-sourcing has actually increased the total Vehicles-Kilometers-Traveled VKT because they attract users from more sustainable modes and add empty kilometers between trips (Diao et al., 2021; Henao and Marshall, 2019; Tirachini and Gomez-Lobo, 2019; Wu and MacKenzie, 2021; Shi et al., 2021; Choi et al., 2022; Agarwal et al., 2023), thus worsening the mentioned issues.

In order to leverage the ability of connecting a large number of users and vehicles in real-time, without worsening traffic conditions, on-demand ridepooling (ODRP) has been proposed (Jiao and Ramezani, 2022). Specifically, ODRP is an on-demand system where the vehicles are fully dedicated to the service, in which several users can share rides at the same time. However, its impact on reducing congestion cannot be taken for granted, as ODRP might attract more users from public transportation than from private modes, which could imply an increase in VKT (Ke et al., 2020; Tirachini et al., 2020; Zwick et al., 2021; Fielbaum and Pudāne, 2024). This problem can be faced by not making ODRP *compete* against public transportation (Poinsignon et al., 2022), but *integrating* this technology in order to offer a better public transportation service. This integration can be facilitated in the future with the implementation of multimodal Mobility-as-a-Service (MaaS) platforms that feature a single payment system for multiple modes.

Designing such an integrated system is methodologically challenging. On the one hand, the optimal design of a public transportation network is an NP-Hard problem for many different specifications (Borndörfer et al., 2007; Fielbaum et al., 2018) even when line frequencies are not considered, let alone if some vehicles follow flexible routes. On the other hand, regarding ODRP, deciding how to group users and how to assign them to vehicles at a city scale can become extremely complex, as the number of feasible combinations becomes too large (Alonso-Mora et al., 2017). All of this justifies studying some specific integration approaches: in particular, as detailed in Section 1.2, most research has imposed that ODRP has to operate feeding fixed-line trunk public transportation services, usually in low-demand areas. The rationale behind this imposition rests on two reasons:

1. Some authors have compared fixed against flexible lines, finding that flexibility is better when demand is low; otherwise, too many vehicles are required and detours can get too long (Badia and Jenelius, 2020; Li and Quadrifoglio, 2010; Papanikolaou and Basbas, 2021; Fielbaum et al., 2023; Quadrifoglio and Li, 2009).
2. Feeder services take all users to a common trunk line, where they might take another (large capacity) vehicle. Having a shared destination eases the search of groups whose routes are compatible, making ODRP more efficient (Fielbaum, 2020).

However, this imposition also has drawbacks. A feeder-trunk scheme forces some users to make transfers, with different but significant levels of discomfort depending on transfer conditions (Garcia-Martinez et al., 2018). Moreover, a feeder-truck system is applicable only under specific circumstances, namely where demand is large enough to admit fixed lines, but some areas can be better served through ODRP. Considering a simplified city representation, Fielbaum et al. (2024) show that imposing ODRP to act as a feeder is optimal only for very particular city configurations.

Moreover, note that a feeder-trunk scheme assumes that only one of the two services (feeder-ODRP or trunk-line-based) is available in each area or for each trip. However, the service design does not need to be black or white. It is possible that both alternatives are available in the same place, for instance at different times. If research has shown that it is efficient to provide ODRP in low-demand areas, it makes sense to expect that it can also be more efficient during low-demand periods: actually, in some districts in Munich[1] this idea is already taking place since 2015. Further, it is also possible to run both fleets simultaneously: as suggested by Fielbaum et al. (2023), ODRP can be efficient during periods of high-demand, as it becomes easier to find users with compatible routes. Moreover, if ODRP is being used during low-demand periods, why not using that same fleet during the peak so that fewer large vehicles are needed? Actually, combining vehicles of different size in the same line can be efficient even when considering only a fixed route but accounting for different periods (Jara-Diaz et al., 2020). When both systems are running at the same time, ODRP can focus on those users that receive the worst service, such as those situated far from the bus stops that need to walk the longest distances.

In this paper, we propose a way to integrate ODRP with traditional public transportation that is still arranged according to lines. Namely, we study a system in which, for every single public transportation line, there are both small vehicles following flexible routes, and large vehicles following fixed routes. A user will only use ODRP when it offers better quality of service for her, and if the system is able to allocate her a vehicle, otherwise she will take a bus (we assume equal fare for everybody). The small vehicles are the only ones in service during the low-demand periods, and continue running during the high-demand periods, when they complement the service offered by large vehicles.

Our mixed system does not impose transfers, and it gets constantly adapted to the varying conditions of the demand during the day, so it is more versatile than a fixed line (that does not adapt well to low-demand conditions, such as during the night) and than the feeder-trunk models previously proposed in the literature. As ODRP becomes part of the public transportation system, some of its users will be required to walk to optimized pick-up and drop-off locations, which is usual in public transportation and has been shown to greatly increase the efficiency of ODRP (Fielbaum, 2021; Fielbaum et al., 2021; Wang et al., 2022; Gurumurthy and Kockelman, 2022).

[1] In several towns around Munich, the public transportation agency sends private taxis to provide mobility services when demand is very low, between 17:30 and 5:45 in working days, see https://www.mvv-muenchen.de/mobilitaetsangebote/bedarfsverkehr/mvv-ruftaxi/index.html. Accessed on 15/02/2023.

1.2. Related works

1.2.1. History of demand-responsive public transportation services

The idea of incorporating into public transportation some vehicles that do not follow fixed routes, but adapt them to the demand, is far from being new. Already in the 1970s several on-demand public transportation services were operating in countries like the US and Germany, and researchers were obtaining the first conclusions of which factors could make these systems succeed (Fielding, 1975; Guenther, 1976). This trend continued in the subsequent decades, discussing scheduling problems (Kikuchi, 1984), comparing the performance of on-demand and fixed-line services (Daganzo, 1984), the relevance of technological innovations to make on-demand systems more attractive (Teal, 1993), and the challenges to integrate them into the general public transportation network (see Kirchhoff (1995), who focus on the German case). In the early 2000s, thanks to more powerful computational resources, simulation studies started to become more popular, which were especially useful to account for the unknown demand patterns that an on-demand service would need to fulfill (Quadrifoglio et al., 2008; Li and Quadrifoglio, 2010). Papers prior to 2014, based on agent-based simulators to study demand-responsive public transportation, were surveyed by Ronald et al. (2015).

Precisely around those years, the first app-based on-demand transportation (Uber) platform began to become widespread. This was a game-changer, both in terms of immediacy (reducing response times), massiveness (increasing the number of passengers and vehicles that could be handled at the same time), and efficiency (high passenger-driver matching rates). While ride-sourcing companies like Uber are mostly focused on non-shared private trips, the same technology could also be applied to shared trips, opening many new possibilities for demand-responsive services in public transportation. We will now discuss in more detail the most relevant studies and trends in this arena from recent years. A survey systematizing the state-of-the-art on flexible on-demand public transportation has been conducted by Vansteenwegen et al. (2022).

1.2.2. ODRP as a feeder

In recent years, different types of on-demand services have been proposed as potential complements to enhance public transportation. This includes ride-sourcing (e.g. Ke et al. (2021) and Yan et al. (2019)), where users do not share a vehicle simultaneously, and ridesharing (e.g. Stiglic et al. (2018)), where people driving their own vehicles can pick up other passengers if convenient for everybody. Here we focus on ODRP, because it is more sustainable (rides are shared), and it can be wholly integrated into the public transportation system (vehicles are fully dedicated to this service). Carrese et al. (2023) made a review of different papers integrating *shared automated vehicles* into public transport, i.e., the automated versions of both ride-sourcing and ODRP.

To the best of our knowledge, research about public transportation systems that integrate ODRP has focused on the utilization of ODRP to solve the so-called *first-and-last-mile problem*, i.e., to leverage the flexibility of ODRP to connect exact origins and destinations with public transportation stations.[2] This means that ODRP is assumed to act as a *feeder*, bringing the users towards the *trunk* lines, that are often (but not necessarily) rail-based.[3]

Such feeder-trunk schemes are studied by several authors, including Bürstlein et al. (2021), Shen et al. (2018), Fielbaum (2020), Chen et al. (2020) and Huang et al. (2022). Bürstlein et al. (2021) show that there is great potential to offer a convenient system, if all people that use their cars to reach the public transportation stations would switch to ODRP, similar to the approach by Shen et al. (2018) for users traveling by bus towards metro stations; Fielbaum (2020) focuses on a high-level analysis to obtain explicit equations that govern a feeder ODRP system; Shen et al. (2018) proposes an integrated system tailored for Singapore, considering also the case in which the feeder vehicles are on-demand but not shared; and Huang et al. (2022) go one step further, by coordinating the operation of the ODRP vehicles with the schedule of fixed public transportation lines. A slightly different system has been proposed by Ma et al. (2019), where ODRP can be used either as a feeder or to make the full trips.

Still on the supply side, Calabrò et al. (2023), Maheo et al. (2019) and Pinto et al. (2020) study how to adapt a fixed public transportation network when ODRP is also part of the system. Calabrò et al. (2023) use a continuous-approximation scheme to show that a system where feeder lines can switch from fixed-routes to on-demand vehicles, depending on the hourly conditions, would outperform offering a solution that remains the same during the whole day. Pinto et al. (2020) also allows that some low-demand fixed-lines can become on-demand, so when they design the whole public transportation network, they decide for each line whether it will be fixed-route or on-demand; Maheo et al. (2019) follow a similar approach, where the ODRP system is not door-to-door but based on previously defined bus stops.

All the papers above focus on the supply side, and usually report promising results regarding how efficient such a combined transportation system can be, mostly by diminishing walking (e.g. Calabrò et al. (2023)) or waiting (e.g. Pinto et al. (2020)) times. However, when the potential demand is also analyzed, those conclusions are disputed, as shown by recent real-world pilot projects. Thorhauge et al. (2022) surveyed users of a real-life first-mile connector in the campus of the Technical University of Denmark, finding that ODRP had a “limited effect on the overall choice of mode for the entire trip chain in their current form (...) The first-and-last-mile transport only represents a limited part of the trip and the effect it has on the entire trip chain is thereby limited”. Similarly, Zuniga-Garcia et al. (2022) study the results of two pilot programs in Austin, Texas, to integrate on-demand systems for the first and last miles, finding an overall mild effect on public transportation ridership and that the pilots were “were mainly used for intrazonal door-to-door trips and not for first-and-last-mile trips”. A more successful example in Mexico City, reported by Tirachini et al. (2020), has been specifically designed to cater for relatively long (mostly commuting) trips, with its demand coming from car-based modes but also from fixed-route public transportation. All of this suggests that, in order to unleash the full potential of ODRP as a public transportation complement, it is crucial not to restrict its use to solve the first and last mile problems.

[2] In their survey, Carrese et al. (2023) also mentions only this type of integration, as well as fully replacing fixed lines with ODRP.

[3] Interestingly, this situation changes when active modes are also considered (typically walking, shared bikes, or biking infrastructure), as some studies do consider ODRP playing a different role when integrated with, or compared to, active modes (Bruzzone et al., 2021; Ferretto et al., 2021; Ma and Jin, 2024).

1.2.3. Splitting public transportation demand

In the conventional feeder-trunk case, ODRP and the fixed lines do not coexist in the same space. In other words, it is assumed that every stage of a trip can only be done in one mode. When both modes are operated jointly and either of them may be used for the same trip, some users will travel via ODRP and others via fixed-line buses, a decision that might be taken either by the users or centrally by the system.

To the best of our knowledge, no previous paper has formally analyzed the situation of co-existence and choice between ODRP and fixed-route public transportation with a theoretical and simulation model, that considers the deployment of both modes altogether. A related problem that has received attention in the literature is the coexistence of two conventional public transportation modes (e.g., bus and rail services as substitutes for the same trip), which is of course a very common situation in many countries. In this case, Hörcher and Graham (2022) have theoretically shown that coexisting modes can be beneficial if users have strong heterogeneity (so that both modes will be preferred by some users), and if scale economies are not too strong (because scale economies are not fully leveraged when the demand is split). Another contribution by Gronau (2000) studies a system in which a public transportation line is divided into two tiers of service following the same route, differentiated by price and quality. On a related note, some authors have studied whether users perceive different public transportation modes similarly. A classic study by Ben-Akiva and Morikawa (2002) shows that bus-based and rail-based services are perceived similarly if they have comparable quality of service, but a bias favors rail when it is already better; their model has been extended by Varela et al. (2018), who find that everything else being equal, users in Sweden usually prefer train over bus or metro.

1.3. Contribution and structure of the paper

The main contribution of this paper is to propose a novel way to integrate ODRP into a public transportation network. Instead of the widely studied feeder-trunk scheme, we utilize some ODRP vehicles to run in parallel to buses that follow fixed lines. That is, for any given fixed line, we assume that some passengers will use buses and others ODRP vehicles. To the best of our knowledge, no previous work has studied a method in which the two modes co-exist in this way. This is particularly relevant in urban or high-demand areas, where fixed lines still play a crucial role. As we show throughout the paper, this idea has promising results, and might be extended in other directions to determine the best ways to use ODRP in public transportation other than in the suburbs.

We leverage the versatility of the well-known ODRP assignment method proposed by Alonso-Mora et al. (2017) and its posterior extension by Fielbaum et al. (2021), through a methodological adaption to decide how to divide the users between buses and ODRP. As a result, the users that would receive the worst quality of service in the fixed line have a greater chance of being assigned to ODRP, which leads to a reduction in the fleet size of the fixed line and an improvement in the average quality of service to users. As discussed in Section 1.2.3, the idea of splitting users among different modes has been studied before, but not when the two involved modes are ODRP and fixed lines. Crucially, we propose a splitting method that is centrally decided, but that presents incentive compatibility so that users are content with their assignment. Instead of differentiating by price (like Gronau, 2000), which raises equity issues, we prioritize users with bad quality of service.

In their literature review, Vansteenwegen et al. (2022) argue that a frequent shortcoming of recent literature about on-demand mobility applied to public transportation is their "lack of insights". As they argue, in most papers the main contribution is the algorithm itself, so that little can be learned in terms of how and when it is convenient to use ODRP integrated in a public transportation service. In this paper, we do provide three strategic insights that can be used for other models as well: (i) We show that ODRP can be used to complement public transportation in high-demand urban zones (without relying on a feeder-trunk rationale), (ii) operating both systems in a common area is a promising strategy, as long as there is an efficient way to split the passenger demand, and (iii) The concrete idea we propose, of ODRP running in parallel to fixed lines, works very well under the conditions analyzed.

In the literature, it is common that ODRP systems are tested with data from one city only, not being possible to compare the performance of the proposed solution across different contexts. As a second contribution, we analyze the merits of our integration strategy in four real-life scenarios from two cities: Santiago, Chile, and Berlin, Germany, with significant differences in demand configurations, operator costs and values of travel time savings. Our simulations do show that the local context matters when it comes to quantifying the users' and operators' cost savings that are reachable when introducing ODRP. We find that: (i) The larger bus operator costs in Berlin, relative to Santiago, and the larger crowding levels that are acceptable in Santiago, make cost savings greater in Berlin when implementing an ODPR system that complements buses; (ii) Low-demand lines benefit more from being complemented by ODRP; and (iii) When the demand is concentrated, the benefits of the ODRP complement are greater. Therefore, our simulations illustrate how relevant local contexts are for the relative convenience of this type of mixed fixed-flexible transportation strategy.

We compare scenarios with human-driven and automated vehicles (that allow to save a fraction of the operator cost). On the one hand, if vehicles are automated, we estimate that in all cases it is possible to offer a better quality of service with reduced operators' costs (a reduction of both operator costs and users costs is what we call a Pareto improvement situation). On the other hand, if vehicles are human-driven, operators' costs likely increase with the mixed strategy, but this cost increase is more than compensated by a reduction in users' cost, in a way that the sum of users' and operators' costs still diminish significantly.

The rest of the paper is structured as follows. Section 2 explains the mixed system in detail, and the methodology to compute its operation; Section 3 presents the real-life scenarios where we simulate the system and the corresponding results; and Section 5 summarizes our findings and presents directions for further research.

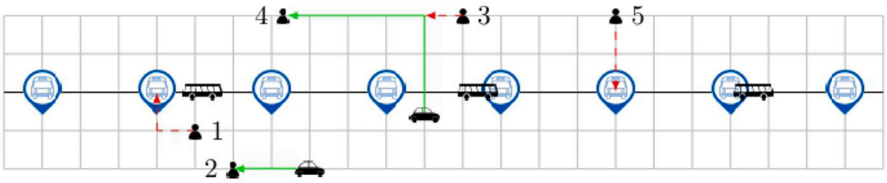


Fig. 1. Simplified representation of the mixed system where the users can perform their trips either via the traditional fixed line, or via the flexible vehicles. Five users need to travel: users 2, 3 and 4 are served by ODRP, where user 3 needs to walk to a pick-up location different than her origin and will share the vehicle with user 4. Users 1 and 5 walk to their closest bus stop.

2. Methodology

As our proposed mixed system is still arranged by lines, the methodological analysis can be done for each line separately as we now do. Consider an area served by a public transportation line, represented by a directed graph $G = (V, E)$. Each edge $e \in E$ requires a time $t_v(e)$ to be crossed in a vehicle, and $t_w(e)$ when walking, which can be done in either direction. We will also denote by $t_v(u_1, u_2)$ and $t_w(u_1, u_2)$ the corresponding times to go from node u_1 to node u_2 when following the shortest paths between them in a vehicle or walking, respectively. The travel demand is represented by a set of users $r \in R$, where each r is a triplette $r = (o_r, d_r, t_r)$ representing the origin, destination, and emerging time (i.e., the time at which the user begins her trip), respectively. The demand is not known beforehand, so request r becomes known to the system at time t_r. The origins and destinations are assumed to be nodes in the graph. The demand is assumed to be fixed, so we can study how efficient our system is in serving it, compared to a traditional one. Understanding whether different users would be attracted by the system would require coupling our analysis with mode choice models, which is beyond the scope of this paper and regarded as future research.

Let us summarize the modeling assumptions we include in order to keep a tractable model. The impact of each of these assumptions is discussed in Section 4.

(A1) We assume that the OD matrix is known, deterministic, and does not change when the service changes. In other words, the set R will be considered exogenous and without any noise. Moreover, when we run numerical simulations, this OD matrix will represent what is currently served in the fixed public transportation line.
(A2) Congestion is disregarded. That is, traveling times $t_v(e)$ do not depend on the number of vehicles in the network or the arc.
(A3) We assume that a single (average) value of time per trip stage is representative of the demand. Below we introduce parameters $\alpha_{Walk}, \alpha_{Wait}, \alpha_{IV}$ to represent the equivalent monetary value of access, waiting and in-vehicle time savings of our travelers.

We consider a mixed model, in which there is a traditional public transportation line and a flexible system operating together. When a user needs to travel, she will request a service and the system will allocate her to a flexible ODRP or to a fixed-route bus, depending on the availability of ODRP vehicles, and on which one provides her with a better quality of service measured via[4] the generalized cost of access, waiting and in-vehicle time; crucially, this decision of the system (whether to allocate her to a bus or to ODRP) is guaranteed to be aligned with the user's interest, as detailed in Section 2.2. The monetary costs incurred by offering the ODRP service can be partially compensated by reducing the fleet required to operate the fixed line, as it will now serve fewer users. Implicit in the description above is the following assumption, whose relevance we also discuss in Section 4:

(A4) All users first request a trip via ODRP, as opposed to some users deciding directly to use the bus without considering ODRP.

Let us remark that nothing would prevent a user from walking to a bus stop and waiting for the bus as in traditional bus systems. Therefore, (A4) is indeed a simplifying assumption.

The ODRP service is not necessarily door-to-door: instead, it is the system that decides where to pick up and drop off every passenger, taking into account both the discomfort induced by walking and the savings that can be gained thanks to avoiding unnecessary detours. This is done following the procedure by Fielbaum et al. (2021) and is described briefly below in this section. It is noteworthy that short walks are a common feature in public transport. The whole system is represented in Fig. 1.

To illustrate the system in a more concrete way, consider the following example. Alice needs to travel from her origin u_1 to her destination u_2, and she knows that the fixed-line is useful for that purpose (i.e., there are bus stops close to u_1 and u_2). Before leaving her origin, she enters the public transportation app, where she might receive one of two outputs: (i) she is told that there are no flexible vehicles available for her, meaning that she needs to take a fixed-route bus, or (ii) she is assigned to a flexible vehicle, and she is told which are the pick-up and drop-off nodes for her trip. We now explain how these decisions are made, and how are the two sub-systems modeled. At the beginning of the following two subsections we provide a table summarizing the notation utilized in each of them.

[4] There are other aspects that could be considered as quality of service. For instance, ODRP might be regarded as offering better quality of service because the vehicles are more comfortable and there is no crowding (Tirachini and del Río, 2019); on the other hand, ODRP is less reliable as traveling conditions depend on the circumstantial co-travelers (Kucharski et al., 2020; Fielbaum and Alonso-Mora, 2020; Alonso-González et al., 2020).

Table 1
Glossary of terms used in the fixed-line model.

Symbol	Meaning
$S_1, \ldots, S_n, T_1, \ldots, T_n$	Bus stops
Per_i	Period i
$Freq_i$	Bus frequency in period i
Dur_i	Duration of period i
$UC_{Fixed}(r)$	Cost faced by user r if traveling by bus
$\alpha_{Walk}, \alpha_{Wait}, \alpha_{IV}$	Value of access, waiting, and in-vehicle time savings
$Walk(r)$	Walking time faced by r if traveling by bus
$Wait(r)$	Waiting time faced by r if traveling by bus
$IV(r)$	In-vehicle time faced by r if traveling by bus
$S_{in}(r), S_{out}(r)$	Stops where user r boards and alights a bus
$Q(r)$	In-vehicle time spent by r waiting at bus stops
t_{BOA}	Time spent by the bus breaking, opening the doors, and accelerating
$N_{Stops}(r)$	Number of stops toured by r while in the bus
$\ell(r)$	In-vehicle time spent by r waiting for other users to board or alight
t_0	Time needed by each user to board or alight
$y(r)$	Number of users boarding or alighting while r is in the bus
OC_{Fixed}	Fixed-line operators' costs
$Fleet_p$	Fleet used during period p
$Q_O(p)$	Av. time spent by buses at each stop during period p
$cycle_p$	Buses' cycle time in period p
$c_{BC}, c_{KC}, c_{BO}, c_{KO}$	Parameters defining the operators' costs
K	Capacity of the vehicle

2.1. The fixed line

The fixed public transportation line is spatially represented by a set of stops $S_1, \ldots, S_n, T_1, \ldots, T_n$. Stops S_i represent the way forth (let us call it direction S) and T_i the way back (direction T).[5] Each stop corresponds to a node in the graph, so that the vehicle visits them sequentially, following the shortest path between two consecutive stops, going to T_1 after visiting S_n and to S_1 after T_n. Further, the day is divided into periods $Per_1, \ldots, Per_m$ (exogenously defined), so that the frequency of service (veh/h) offered by the public transportation agency is fixed within those periods, i.e., the number of buses touring the circuit of stops depends on the period; for the sake of simplicity, we do not model explicitly the transition between consecutive periods. We denote $Freq_1, \ldots, Freq_m$ the corresponding frequencies, and $Dur_1, \ldots, Dur_m$ the duration of each period. Our objective is to write users' and operators' costs as a function of $Freq_1, \ldots, Freq_m$. A summary of the notation utilized in this subsection is provided in Table 1.

Users' costs. The costs of user r depend on average walking, waiting, and in-vehicle times:

$$UC_{Fixed}(r) = \alpha_{Walk} Walk(r) + \alpha_{Wait} Wait(r) + \alpha_{IV} IV(r), \tag{1}$$

where the parameters α represent the corresponding values of access, waiting and in-vehicle time savings. We now explain how to calculate each of these components. Recall that we assume equal fares for everybody, which is why they can be omitted in Eq. (1).

Consider a user r taking the fixed line, that we assume (without loss of generality) traveling in direction S. She will select the stops $S_{in}(r), S_{out}(r)$ that minimize total walk, which is then given by

$$Walk(r) = t_w(S_{in}(r), o_r) + t_w(S_{out}(r), d_r) \tag{2}$$

The waiting time faced by r depends on the frequency of the public transportation line. Defining $Per(r)$ as the period corresponding to t_r, then if bus headways are constant, passengers arrive randomly at a constant rate and vehicle capacity constraints are not binding, r will wait half of the headway on average

$$Wait(r) = \frac{1}{2Freq_{Per(r)}} \tag{3}$$

Finally, the in-vehicle time (i.e. the time spent on the vehicle) depends on the riding time between bus stops, and on the time spent at each bus stop:

$$IV(r) = \sum_{j=in}^{j=out-1} t_V(S_j, S_{j+1}) + Q(r), \tag{4}$$

[5] The extension to consider different number of stops per direction, or to consider a circuit instead of two directions, is straightforward, but this scheme eases the rest of the description.

where the first term is the time spent in-motion, and $Q(r)$ is the time spent at bus stops, which depends on r not only because longer trips visit more bus stops, but also through $Per(r)$, as high-demand periods face a greater load per bus, inducing longer times at the stops due to boarding and alighting of passengers. To be precise:

$$Q(r) = t_{BOA} \cdot N_{Stops}(r) + \ell(r) \tag{5}$$

In Eq. (5), the first addend represents the time spent breaking, opening the doors, and accelerating, so t_{BOA} is the time spent per stop (independent of r) and $N_{Stops}(r)$ is the number of stops visited by r, which depends on r but not on the frequencies. The second addend $\ell(r)$ represents the time waiting for other passengers to board and alight, calculated as

$$\ell(r) = t_0 \cdot y(r) \tag{6}$$

where t_0 is the time needed for each user to board or alight, and $y(r)$ is the sum of the number of users boarding or alighting at each of the N_{Stops} stops visited by r. The term $y(r)$ can be directly computed by dividing the period $Per(r)$, that lasts $Dur_{p(r)}$, into $\lceil Dur_{Per(r)}/Freq_{Per(r)} \rceil$ sub-intervals, representing the different headways in which users accumulate at each bus stop. As we know the time at which each passenger arrives at the bus stop, it is straightforward to determine which passengers share the same bus with r.

Operators' costs. Total operators' costs encompass fixed and usage-dependent costs. It has been estimated with data from several countries (e.g., Sweden, Australia, Germany, Chile) that both cost components grow in an affine linear way with respect to the size of the vehicles utilized (Jansson, 1980; Tirachini and Hensher, 2011; Tirachini and Antoniou, 2020). That is to say, some cost items are constant regardless of the vehicle size (e.g., drivers' salaries), and other cost components increase for larger vehicles (e.g., energy consumption). Similarly, some costs do not depend on vehicles' usage (e.g., capital costs) and others do (e.g., again, energy consumption).

Let us denote by $cycle_p$ the cycle time required to tour the whole circuit during period p, which is given by the time required to travel between all consecutive bus stops plus $2nQ_O(p)$, corresponding to the time lost at each stop, where we denote by $Q_O(p)$ the average time spent per stop during period p (the sub-index O refers to operators). In general, frequency is given by the fleet divided by the cycle time, thus the fleet utilized during period p can be computed as $Fleet_p = Freq_p \cdot cycle_p$, leading to total operators' costs:

$$OC_{Fixed} = \max_{p=1,\dots,m} Fleet_p \cdot (c_{BC} + K c_{KC}) \tag{7}$$

$$+ \sum_{p=1}^{m} Fleet_p \cdot Dur_p \cdot (c_{BO} + K c_{KO}),$$

where the first term corresponds to fixed costs (through the fixed parameter per bus c_{BC}, and the slope c_{KC}) and the second term to operating costs (with parameters c_{BO}, c_{KO}). In both terms the cost grows linearly with the capacity K of vehicles (measured in passengers per vehicle). For a given vector of frequencies, we compute the maximum number of users that would be on the same vehicle at the same time: this can be done straightforwardly as we know, for each user, in which stop and headway she boards the vehicle, and where she alights. We define K to be equal to that value.

We consider as the objective function of the system its total costs, i.e., the sum of users' and operators' costs. Other objectives and performance measures could be considered (e.g. VKT, fleet size, or quality of service), but the total costs are more general as they capture every aspect of the system (Jara-Díaz, 2007). Note that both users and operators costs depend solely on the vector of frequencies-per-period. In our modeling approach the system can be simulated with the optimized values of frequencies, or these values can be taken as exogenous, for instance, taking into account the fleet configuration of the current public transportation system in a city. When we run simulations (most of Section 3), we assume optimal frequencies for most of our results. We do include an additional case considering exogenous real-life frequencies at the end of Section 3.

2.2. ODRP designed to complement fixed lines

We now explain the operation of the ODRP sub-system. A summary of all the notation utilized in this subsection is provided in Table 2. As in the fixed line, users' costs are defined through Eq. (1). Operators' costs are also the sum of fixed and usage-dependent costs, as in Eq. (7); however, the systems operates in a completely different way, as the vehicle routes are not fixed but respond to the demand. Whereas the usage-dependent costs in Eq. (7) depend on the number of vehicles per period, in ODRP all vehicles are active during the whole day, but they may remain idle (thus not contributing to usage-dependent costs) for some time when they have no requests assigned. The resulting costs for users and operators are computed through simulations that we now explain.

As this system operates on-demand, the main question is how to assign users to vehicles. To do this, we adapt the method by Fielbaum et al. (2021), an extension of the well-known assignment technique from Alonso-Mora et al. (2017) to optimize the pick-up and drop-off points of each passenger. Flexible vehicles are usually not enough to serve everybody, and those users that are not served by ODRP will use the fixed lines. Hence, there are two aspects that we need to account for:

- If a user will receive a worse quality of service in ODRP than with the fixed lines, she should not be assigned to ODRP (otherwise she might not accept the assignment and use the fixed line anyway).

Table 2
Glossary of terms used in the ODRP model.

Symbol	Meaning
Δ t	Lapse of time to accumulate a batch of requests
R_t	Set of requests to assign at time t
$req(T)$	Requests forming potential trip T
$veh(T)$	Vehicle serving potential trip T
κ	Capacity of the vehicles
$\mathcal{T}$	Set of trips
$\mathcal{V}$	Set of vehicles
$UC(T)$	Users' costs yielded by trip T
$UC_T(r)$	Cost faced by user r if traveling through trip T
$\Delta UC_T(r)$	Extra cost faced by user r if trip T takes place
$Pax(v)$	Passengers currently being served by vehicle v
x_T	Binary variable marking whether trip T is taking place
z_r	Binary variable marking whether request r has to take a bus
$\psi(r)$	Penalty of non-serving r via ODRP

- Not all users face the same conditions in the fixed line. Therefore, when deciding whom to serve via ODRP, we aim to include this criterion and give a greater chance to the ones receiving the worst quality of service in the fixed line.

We now briefly explain how the assignment method works, and how do we capture these two aspects. Assignments are decided in a *receding horizon* fashion, i.e., requests are accumulated during Δt (we use $\Delta t = 1$ min) and assigned all together, which is done iteratively.

Let us denote R_t the set of requests to be assigned at time t. The assignment method consists in two steps:

1. First, the set of feasible ***trips*** is computed, where a trip T is defined by a set of requests $req(T) \subseteq R_t$, together with a vehicle $veh(T)$ (that might be serving previous requests), pick-up and drop-off points for each request in $req(T)$, and an updated route for $veh(T)$. The pick-up and drop-off points, as well as the updated vehicle's route, are decided optimally[6] according to Eq. (8) below. For a trip to be feasible, the capacity κ of the vehicle cannot be exceeded (here we use $\kappa = 4$) and hard constraints on waiting time, walking time and total delay must be fulfilled. Let us denote by $\mathcal{T}$ the set of feasible trips, and $\mathcal{V}$ the set of vehicles. Each trip T imposes user costs $UC(T)$, affecting both the users in $req(T)$ and the ones previously being served by $veh(T)$, as their traveling times can increase due to new requests. We denote by $UC_T(r)$ the costs faced by $r \in req(T)$ if traveling through the trip T, and by $\Delta UC_T(r)$ the extra costs faced by $r \in Pax(veh(T))$ if trip T is executed, where $Pax(v)$ are the users currently being served by vehicle v, thus
$$UC(T) = \sum_{r \in req(T)} UC_T(r) + \sum_{r \in Pax(veh(T))} \Delta UC_T(r) \tag{8}$$
Note that we use $UC(T)$, meaning that the operators' costs are not taken into account at the assignment step. Such costs, which are proportional to the length of the resulting route, would bias the system towards serving via ODRP those passengers traveling short distances; doing so might make sense if ODRP was running alone (which is why Fielbaum et al., 2021 do include operating costs when assigning), but not in this context, as then the buses would receive a higher load, increasing their costs. Therefore, it is better not to include any bias with respect to the length of the trips. For the same reason, in-vehicle times only consider the extra-time over the length of the shortest path between origins and destinations.[7]
Computing $\mathcal{T}$ can be computationally expensive. This is faced by leveraging the following property as explained by Alonso-Mora et al. (2017): for T to be feasible, a necessary condition is that $(H, veh(T))$ is also feasible for every $H \subset req(T)$. We omit here, and include in the Appendix A, those details of the method that are already described by Fielbaum et al. (2021).
2. Second, the following ILP is solved:
$$\min_{x,z \in \{0,1\}} C(x,z) = \sum_{T \in \mathcal{T}} x_T UC(T) + \sum_{r \in R} z_r \psi(r) \tag{9}$$
$$\text{s.t. } z_r + \sum_{T: r \in req(T)} x_T = 1 \forall r \in R_t \tag{10}$$
$$\sum_{T: veh(T) = v} x_T \leq 1 \forall v \in \mathcal{V} \tag{11}$$
where binary variables x_T represent which trips are selected to be executed, whereas variables z_r mark which requests are not served by ODRP and must use the fixed line instead. Eq. (9) is the objective function, which includes the users costs for

[6] As explained in the Appendix A, some heuristics might be applied if the scale of the problem is too large.

[7] We remark that this is a heuristic decision, as splitting the passengers between ODRP and public transportation in a way that minimizes total costs (or more generally, that maximizes social welfare) is an extremely complex problem. The complexity of finding an optimal solution is actually greater because the ODRP assignment problem involves several NP-Hard problems (Fielbaum et al., 2021) and is a dynamic one, so ideally one should consider the future demand when deciding how to assign the current demand.

each trip and a so-called “non-service penalty” $\psi(r)$ for each unserved user r. As discussed below, this penalty will play a key role to decide which users to serve through ODRP. Eq. (10) ensures that each request is either served by ODRP or marked with z_r, and Eq. (11) means that each vehicle can be assigned to at most one trip at a time.

This ILP is solved using a state-of-the-art commercial solver (Gurobi). For all four scenarios we study in Section 3, all of them based on real-life public transportation lines with their corresponding demand, it is possible to solve the ILP to optimality in less than a minute for every assignment step.

In the original method (Fielbaum et al., 2021), a constant high value $\psi(r)$ is assumed for every request, so that there is no prioritization among users and they are only non-served when it is unfeasible or very inefficient to serve them. Instead, here we define the non-service penalty equal to the expected cost that would be faced by the user if taking the fixed line:

$$\psi(r) = UC_{Fixed}(r) \tag{12}$$

Eq. (12) modifies the assignment decisions to solve the two issues described above. First, the presence of $\psi(r)$ in the objective function (Eq. (9)) means that, when assigning a vehicle v to one of two users r_1, r_2 because it is not possible to serve everybody, one criterion to decide which user to serve will be which is the one having the worst (more costly) quality of service in the alternative fixed-line service, which encompasses all travel time stages experienced by the user, i.e., walking, waiting, and in-vehicle times. Note that this criterion is to be combined, through $UC(T)$, with how efficient is v in serving each request, namely how close is v to r_1 and r_2, which is related to the resulting waiting times, and the required deviation on the current route of v, as this affects the users currently on-board (if any).

The other issue to be faced is that no passenger should be assigned to ODRP if it offers her a worse quality of service than the fixed-line. The fact that this is solved is formalized in the following proposition.

Proposition 1. *Let $T \in \mathcal{T}$ such that the optimal solution (x, z) of the ILP has $x_T = 1$ (i.e., that the users in $req(T)$ are assigned to $veh(T)$). Then $\forall r \in req(T), UC_T(r) \leq UC_{Fixed}(r)$.*

Proof. In plain words, the proof consists of showing that, if there was an r such that $UC_T(r) > UC_{Fixed}(r)$, then removing r from T, so that r travels by bus and everything else remains unaffected, would be a better solution than (x, z). This is true because r would be better (which is the exact meaning of $UC_T(r) > UC_{Fixed}(r)$), and the other users served by the same vehicle would also improve as they can now follow a shorter route. The mathematical formalization of this proof can be found in the Appendix A. □

The corollary of Proposition 1 is that **the system's decisions are aligned with users' interests**. In other words, a user r has no incentive to deviate from the system's decision: if the system tells r to take a bus, she has no other choice; on the other hand, if the system assigns r to a flexible vehicle, she could reject the assignment and take a bus instead, but this would be harmful for her as Proposition 1 guarantees that the ODRP service is better.

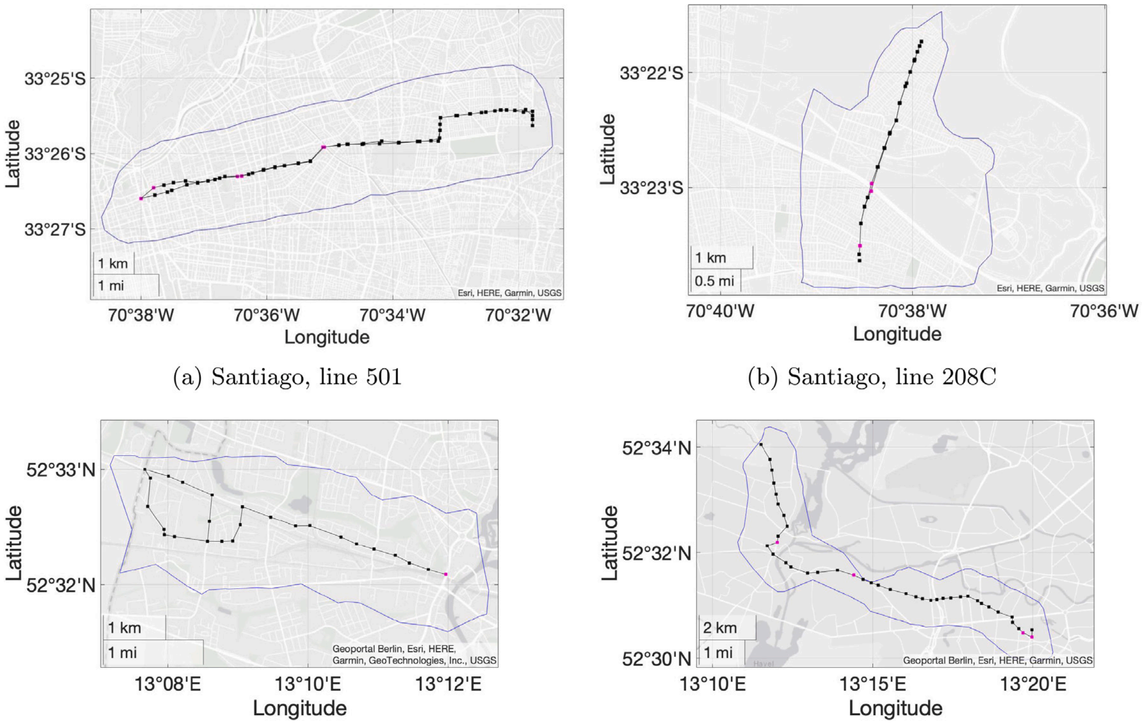
This corollary highlights a relevant virtue of our system. As decisions are taken dynamically, the users do not have to make a choice with incomplete information. In fact, if the buses and the flexible vehicles would operate in an uncoordinated fashion, every user would need to decide a priori which of the two systems to take depending on the expected quality of service of each of them (and, if they are risk averse, considering that ODRP is usually more unreliable, as argued by Kucharski et al., 2020; Alonso-González et al., 2020; Fielbaum and Alonso-Mora, 2020). Instead, the assignment method we use is better for the users, as they know the decision is aligned with their own interests and taken with more information about the ODRP system — using the actual cost rather than the expected one; and is also better at a system-wide level, as the decisions are taken in a centralized way (the problem of decentralized decisions is thoroughly analyzed by Bahamonde-Birke (2022), who argues that it could even lead to Braess-paradox-type of situations).

Rebalancing. After deciding the assignment, a rebalancing step is performed. This is done exactly as proposed by Alonso-Mora et al. (2017), namely sending idle vehicles (if any) towards the origins of the requests r marked with $z_r = 1$.

2.3. Adapting the fixed line

When ODRP is included, the number of passengers using the fixed-line system diminishes, so the vector of frequencies per period can be reduced. Operators' costs are expected to increase due to the use of a larger fleet of small vehicles and buses (operators' costs are reduced with a small fleet of large vehicles, see Jara-Díaz and Gschwender (2009)), so we reduce the bus frequency per period as much as possible, without violating the bus capacity constraint: these frequencies determine the fixed-line operator costs that we consider when calculating the costs of the mixed system. This has the unavoidable cost of increasing waiting times for the users that remain as bus riders. A corollary of the frequency reduction is that it is not fair to leave, in a given period, a very small number of users in the traditional line, because they would face a very low frequency; to prevent this problem, we serve everybody with ODRP in the periods where it is possible to do so.

It must be also noted that as we assume that all the users are considered as potential ODRP passengers, we need to consider that the ones remaining in the fixed line wait to receive a notification that they are not served in ODRP. As we decide the assignments every Δt, fixed-line users wait on average $\Delta t/2$ (in addition to the time waiting at the bus stop). This extra waiting time shows that our assumption that all the line's passengers are potential ODRP users is actually pessimistic: a more sophisticated demand model, which is out of the scope of this paper, would probably suggest that those users that are well-served by the fixed line would not call a flexible vehicle, because their chances of getting served by ODRP would be low. This would reduce the mentioned extra waiting time, enhancing our results even more.



(a) Santiago, line 501 (b) Santiago, line 208C

(c) Berlin, line 237 (d) Berlin, line M45

Fig. 2. The route and bus stops of Lines 501 and 208C in Santiago, and lines 237 and M45 in Berlin. Purple nodes represent intersections with metro lines. The blue boundary shows the areas served by the lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3. Results and discussion

3.1. The scenarios and data

We study how this mixed flexible-fixed public transportation service would operate during a whole day in Santiago, Chile, in the areas currently served by Lines 501 and 208C, and in Berlin, Germany, for Lines 237 and M45, depicted in Fig. 2. Thus, we cover a middle-income and a high-income country, which have differences in operator cost parameters, modal shares, and values of travel time savings. We consider lines with different demand patterns and levels as detailed below. None of these lines run in parallel to a rail-based mode (they do intersect perpendicular metro lines), so that mode choice within the public transportation system (bus/rail) does not play a role.

The corresponding graphs are composed by the nodes that are no further than 1 km from the bus line, and from which people can reach a bus stop in no more than 20 walking minutes, which are realistic upper bounds for what most users do walk in public transportation systems (Daniels and Mulley, 2013; Durán-Hormazábal and Tirachini, 2016). The boundaries of the networks are shown in blue in Fig. 2.

For Santiago, the demand datasets are computed following the method by Munizaga and Palma (2012), which is based on smartcard fare payment data from the city's public transportation system. The demand per working day in Lines 501 and 208C are 12,250 and 1242 trips/day, respectively, representing cases of high and low demand. The lines intersect some metro stations, shown in purple in Fig. 2. Line 208C is an auxiliary line that only runs during peak periods. In the case of Berlin, the demand level is 16,810 trips/day for Line 237 and 33,000 trips/day for Line M45, and are computed through MATSim simulations (Ziemke et al., 2019). All datasets contain, for each passenger, the stops where she begins and finishes her trip, so we assign the exact origins and destinations randomly within the nodes situated close to such a stop, where the closer nodes have a greater chance to be assigned, to reflect the fact that there tend to be more origins and destinations near transit stops (Kim and Li, 2021). The method for the estimation of trip origins and destinations is explained in detail in the Appendix A.

The edges and nodes for all areas were obtained through OpenStreetsMaps. We begin analyzing the cases with both automated (driverless) and human-driven vehicles, which only differ in the operators costs' parameters. We use specific operator cost parameters for human-driven and automated vehicles for both Chile and Germany, based on the estimations for electric vehicles from Tirachini and Antoniou (2020), and as a result, operators' costs are larger in Germany than in Chile. Following Tirachini and Antoniou (2020), our model also takes into account that larger crowding levels are acceptable in Santiago than in Berlin: based on observed values,

Table 3
Parameters that are different between automated and human-driven vehicles.

Parameter	Value in Santiago	Value in Berlin
Automated vehicles		
c_{BC}	30.28 [$US]	24.6 [$US]
c_{KC}	1.1 [$US \pax]	2.1 [$US \pax]
Human-driven vehicles		
c_{BC}	78.9 [$US]	157.8 [$US]
c_{KC}	1.2 [$US \pax]	2.4 [$US \pax]

bus capacity for a 12-m long bus is set as 70 passengers in Berlin and 90 passengers in Santiago, and for a 18-m long (articulated) bus capacity is set as 110 passengers in Berlin and 140 passengers in Santiago. Crucially, this difference in carrying capacity in the two scenarios will have an effect on the cost convenience of introducing an ODRP system, as our simulations show.

Automated and human-driven vehicles only differ in the fixed costs, i.e., parameters c_{BC} and c_{KC}. The numerical values of these parameters are shown and explained in Table 3. The remaining parameters are described in Table 6 in Appendix A. Note that the largest differences between automated and human driven vehicles occur in c_{BC}, i.e., precisely the term that contains the driver's salary, which can be reduced by more than half thanks to automation.[8]

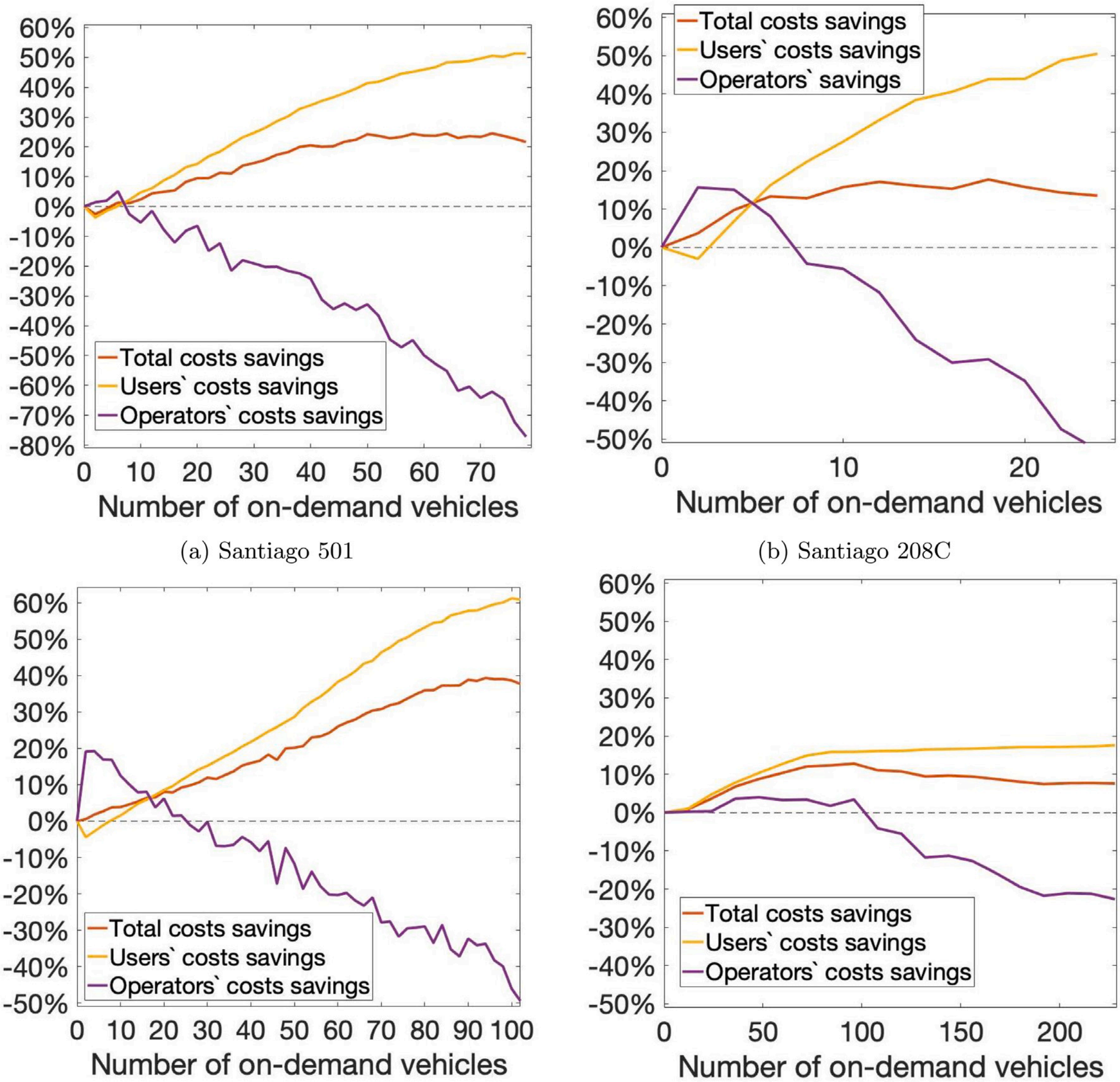
A visualization of the system can be seen in the video attached, for Line 501 in Santiago (see Appendix B).

3.2. Results of the simulations

We simulate the operation of the mixed service during the whole day, and for different numbers of ODRP vehicles. Figs. 3–4 compare the resulting costs of the traditional system (only having the fixed line) against the mixed system that we propose, by exhibiting the savings achieved by users and operators depending on the ODRP fleet. Note that as we are showing the savings, being above the 0% horizontal line means that the corresponding actor is being benefited (users, operators, or overall). Fig. 3 shows the results for automated vehicles, and Fig. 4 for human-driven. Some remarkable conclusions emerge:

- Users are better off under the mixed system with ODRP operating in all four scenarios, regardless of whether vehicles are automated. The only exception is when too few ODRP vehicles are operated (explained below). The superiority in quality of service from ODRP is mainly reached due to a reduced amount of walking.
- If the vehicles are automated, there are segments when the three curves are above 0%, i.e., where both users and operators are benefited. That is, under some circumstances **our system is able to induce a Pareto improvement, in which both users and operators are better-off.** This happens partly due to the good quality of service provided by ODRP, that outweights the extra waiting time experienced by the fixed-line users, but also because buses do not always run full during off-peak periods (Jara-Diaz et al., 2020), so that there is room to reduce the service frequency.
- If vehicles are human driven, a (mild) Pareto-improvement is still possible in three out of the four scenarios (Santiago-501, Santiago-208C and Berlin-237), whereas in the other case (Berlin-M45) either users or operators would end up worse-off. Overall costs can still diminish up to about 10% in the lines we simulate.
  The difference between automated and human-driven vehicles reflects that our model is particularly sensitive to the relative value of the operators' cost parameters. In fact, as the combined system introduces small vehicles and removes large ones, if capacity-independent vehicle costs c_{BC}, c_{BO} become larger compared to the marginal costs of extra capacity c_{KC}, c_{KO}, the performance of our mixed system would become worse. This is exactly the case when vehicles are not automated, as drivers' salary do not depend on the size of the vehicle K.
  In all, our results show that even **with human-driven vehicles, the ability of diminishing walking times can reduce total costs by a significant share; however, this requires allocating extra monetary budget**. As such, vehicles' automation might become a crucial enabler for the system we propose here. This finding coincides with previous papers that show that the competitiveness of on-demand mobility strongly increases with this technology (Oke et al., 2020; Basu et al., 2018; Fielbaum et al., 2023).
- When the number of vehicles is too low, users can become worse off. The explanation for this result comes from what was discussed in Section 2.3: users that remain in the buses now face an average extra 30 s of waiting, namely until the system notifies them that there is no flexible vehicle available. If there are just a few vehicles, almost every user remains in the buses and are thus negatively affected in this way, while the number of users who benefit from the presence of ODRP is smaller.
- Total costs savings reach their maximum but then start to decrease, implying that serving all of the demand with only ODRP would be less efficient than the mixed system. Such an outcome stems from the increased operators' costs when operating only ODRP for large demand levels.

[8] These parameters are derived from Tirachini and Antoniou (2020), who estimate the different components of the cost function (such as maintenance, energy, and capital) for $K = 5, 8, 50, 90, 140$, in Santiago, Chile, and Munich, Germany, for human-driven and automated electric vehicles. We assume the values in Munich to be valid for Berlin as well. We classify each of such components to be either dependent on the number of driving hours (hence contributing to c_{KO} and c_{BO}) or not (contributing to c_{KC} and c_{BC}). Numbers are normalized to account for the number of driving hours per day. After this procedure, we have total fixed and usage-dependent costs for each $K = 5, 8, 50, 90, 140$, so that a linear regression leads to the parameters shown in Tables 3 and 6.



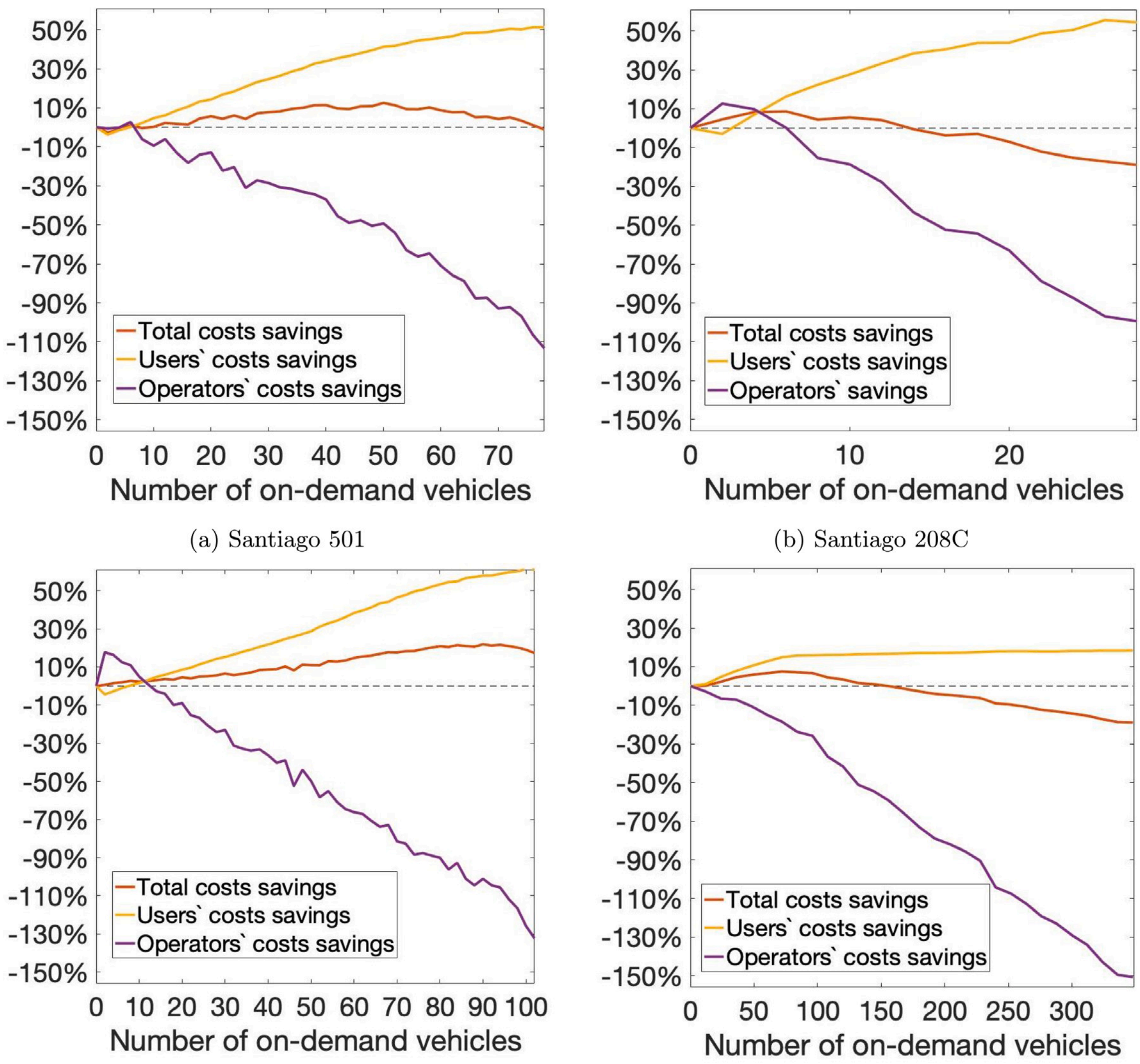
(a) Santiago 501 (b) Santiago 208C

(c) Berlin 237 (d) Berlin M45

Fig. 3. Percentage of the costs saved thanks to utilizing the mixed system with automated vehicles, compared to the original case with no flexible vehicles.

Comparison of the four scenarios. In order to gain more insights from our model, we provide a more detailed comparison of the four scenarios simulated, assuming an operation with automated vehicles, as this technology would yield the largest benefit. While in all of the scenarios it is possible to find Pareto improvements if vehicles are automated (or to reduce total costs if vehicles are human-driven), the magnitude of the changes presents significant variations. To analyze the causes of these differences, in Table 4 we present relevant characteristics of each scenario:

- The total daily demand changes from little more than a thousand (208C in Santiago) to 33,000 (M45 in Berlin). In general, we observe that the scenarios with lower demand levels present better results. The reason behind this is that the fixed routes provide a worse quality of service in low-demand areas, as low frequencies yield large waiting times. Therefore, shifting some trips to ODRP is more beneficial in that context. This conclusion is reinforced by an additional analysis done on Line 237 in Berlin and presented in the Appendix A, where we show that if the demand was roughly half of its true level, the results become significantly better.
- In Santiago, large buses are cheaper to buy and operate (Tirachini and Antoniou, 2020) and run more crowded, so replacing them with smaller vehicles involves larger operators' costs. This is formalized in the third and fourth columns of Table 4, that show that the ratios c_{BC}/c_{KC} and c_{BO}/c_{KO} are larger in Santiago. This last aspect would be nuanced if our model accounted for



(a) Santiago 501 (b) Santiago 208C

(c) Berlin 237 (d) Berlin M45

Fig. 4. Percentage of the costs saved thanks to utilizing the mixed system with human-driven vehicles, compared to the original case with no flexible vehicles.

the crowding discomfort as increasing the value of in-vehicle time savings (as Tirachini et al., 2014; Jara-Díaz and Gschwender, 2003), because every user switching from a crowded bus to ODRP would improve her quality of service even further.[9]

- We report the *Gini index* of each line. To be specific, we calculate how many trips are originated in each bus stop, and compute the Gini index of the resulting OD vector. We remark that the Gini index has been proposed in the past as a way to measure the concentration of trips in public transportation (Hörcher and Graham, 2021). We can observe that the two lines with the greater Gini, i.e., where the demand is more concentrated, present better results. This makes sense, as the fixed lines need to visit the whole circuit regardless of the demand, whereas the flexible vehicles can remain close to where trips are mostly generated.

From a policy perspective, there are two possible objectives one might consider: either to diminish total costs as much as possible, likely requiring more subsidies,[10] or to obtain the maximum Pareto improvement $\mathcal{P}$. This quantity $\mathcal{P}$ is defined as the minimum

[9] Besides crowding, there are many directions in which we could make our bus model could be extended, such as distance-dependent fares (Jara-Diaz et al., 2024), demand-dependent spacing (Lehe and Pandey, 2023), or considering that the time required to alight from a bus depends on its level of occupancy (Seriani et al., 2019). We have opted for a formulation that admits a simple representation of public transportation supply while accounting for the central design elements.

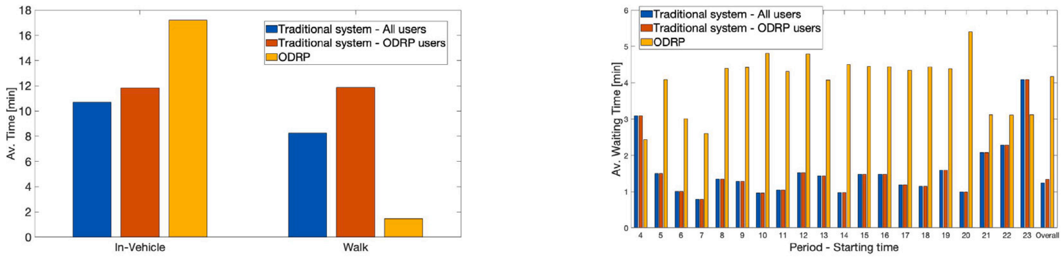
[10] Traditional urban public transportation is usually subsidized (Börjesson et al., 2020; Glaister, 2018).

Table 4
Main characteristics of the four scenarios.

Scenario	Total demand	c_{BC}/c_{KC}	c_{BO}/c_{KO}	Gini
501 (Santiago)	12,250	27.5	26.7	0.44
208C (Santiago)	1242	27.5	26.7	0.66
237 (Berlin)	16,810	11.6	15.3	0.58
M45	33,000	11.6	15.3	0.45

Table 5
Detailed results for each of the four lines, considering either the objective of maximizing total savings, or maximizing the Pareto savings.

Line	Objective	ODRP Fleet	Bus fleet reduction	Total savings	Users savings	Operators savings	ODRP service rate
501 (Santiago)	Max. savings	72	31%	24.5%	50.5%	−62.1%	66.8%
	Max. Pareto	6	11.1%	1.3%	0.2%	5.1%	7.6%
208C (Santiago)	Max. savings	18	28.6%	17.8%	43.9%	−29.2%	71.3%
	Max. Pareto	6	28.6%	13.3%	16.2%	8%	36.7%
237 (Berlin)	Max. savings	94	75%	39.3%	58.9%	−33.7%	90.2%
	Max. Pareto	20	30.6%	8%	8.6%	6.2%	23.3%
M45 (Berlin)	Max. savings	96	28.2%	12.8%	15.9%	3.4%	24.6%
	Max. Pareto	48	15.4%	8.8%	10.4%	4%	17.5%



(a) (b)

Fig. 5. Comparison of service times offered by: the traditional system considering all users (blue columns), the traditional system considering the users that would use ODRP in the mixed system (red columns), and ODRP (yellow columns). Note that the red and yellow columns refer to the same users, comparing the traditional vs the mixed network. Results are shown for the Line 237 in Berlin with 20 flexible vehicles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

improvement between users and operators. To be formal, let us denote by $S_U(F), S_O(F)$ the users' and operators' savings when a fleet of F on-demand vehicles is used. Then the maximum Pareto-improvement is given by

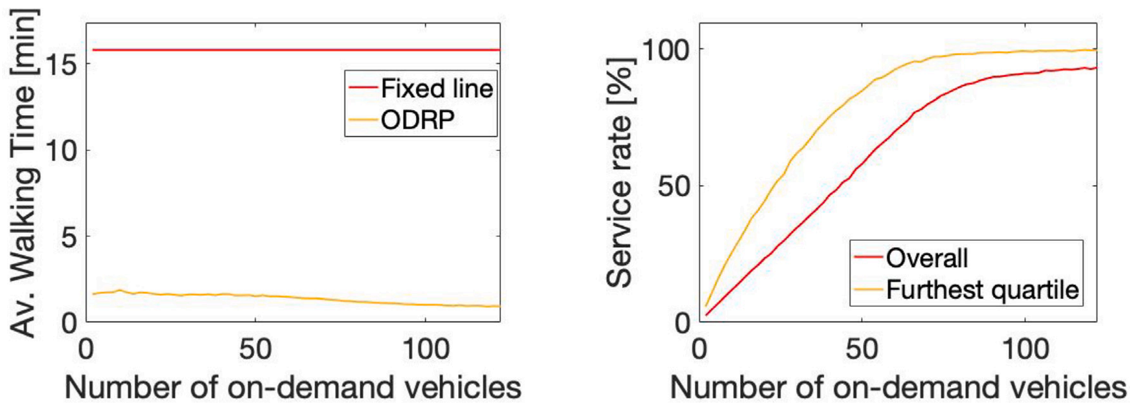
$$\mathcal{P} = \max_F \min\{S_U(F), S_O(F)\}. \tag{13}$$

Intuitively speaking, we are selecting the number of vehicles that benefits both users and operators as much as possible.

In Table 5 we show detailed results regarding the resulting costs, fleets, and service rates through ODRP for both objectives. Crucially, total savings reach from 12.8% (Line M45 in Berlin) to 39.3% (Line 237 in Berlin), while the Pareto improvement is always larger than 4% in all of the lines but 501 (Santiago) where users' savings are mild (0.2%) under this objective.

One scenario in detail. To better understand these results, in Fig. 5 we zoom into the Line 237-Berlin scenario, focusing on the case with 20 vehicles, which presents the largest Pareto improvement (both users and operators improve by more than 6.2%). We analyze the quality of service experienced by ODRP users (yellow columns in Fig. 5), and compare it with what they would have experienced in the system without ODRP (red columns). We also add the average quality of service when there is no ODRP (the baseline — blue columns). As in public transportation the waiting times are strongly dependent on the period, in Fig. 5(b) we depict the results for each period. These figures provide interesting insights (that remain valid in the other scenarios, as shown by the corresponding analogous figures for each of them that can be found in the Appendix A):

- As expected, the average walking time is much lower in ODRP than in traditional public transport, as users receive an almost door-to-door service. The comparison between the blue and red columns for walking in Fig. 5(a) reveals that ODRP serves users that would have had a substantially greater walking time than the average user if there was no ODRP, i.e., ODRP is effectively considering the distance to the public transportation line as a relevant criterion when deciding whom to serve. This finding suggests that **having a significant number of users walking more than 10–15 min in total (access plus egress) is the crucial factor for our system to be better than the traditional line**.



(a) (b)

Fig. 6. Analysis of the advantage of the mixed system considering the 25% users situated furthest from the fixed line 237 in Berlin with 20 flexible vehicles, as the flexible fleet grows: (a) Walking times with or without ODRP, (b) Percentage of users served via ODRP overall, and considering the furthest quartile only. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- In-vehicle times are much smaller in the fixed-line, which happens because it runs straight through the main street. In this line, ODRP serves mostly long trips (the corresponding red column is taller than the blue one), but the contrary happens in other scenarios, which is consistent with the fact that the distance between origin and destination is not a relevant factor to decide which type of vehicle to use.
- In ODRP, waiting times do not change much across periods, and are actually larger during the peak, as more users need to board the same fleet. Therefore, regarding this metric, the users that are favored with the implementation of the mixed system are the ones that use ODRP during the low-demand periods, where the fixed-line frequencies are low as well.
  It is worth recalling that users who remain on the buses now face a greater waiting time, as frequencies are reduced. In our model we assume all users would first request an ODRP vehicle; however, in reality, some users might go directly to bus stops, because of individual preferences or habit, thus never benefiting from the ODRP virtues. In the case under scrutiny (Line 237 in Berlin with 20 vehicles) the average waiting time at the bus stops increases from 1.24 to 1.65 min, i.e., a significant yet definitely tolerable difference.

As the main virtue of the mixed system is its ability to reduce the walking times for some, we now study this effect in more detail. To do so, and still considering Line 237 in Berlin, we focus on the 25% of the passengers situated furthest from the fixed-line (the *furthest quartile*), i.e., the ones that would need to walk more if the ODRP system was not available. In Fig. 6(a) we compare the walking times they would have experienced in the traditional system with the walking times offered by the mixed system, where the advantage of the latter becomes evident. In Fig. 6(b) we compare the percentage of total users traveling in ODRP, with the same percentage considering only the mentioned quartile: as the fleet grows, more users are being served in general, but the ODRP system always serves with a greater probability the ones that would need to walk more, as revealed by the fact that the yellow line lies over the red one. This last conclusion can also be visualized in Fig. 7, where we show for each zone in the network, the fraction of the users transported by ODRP, considering the fleet that achieves the maximum Pareto improvement. The zones are calculated only for this visualization, and are computed following a simple method explained by Wallar et al. (2018). The results clearly show that those zones placed further from the bus line are greener, i.e., they use ODRP more intensely.

Let us discuss these last results through the lens proposed by Hörcher and Graham (2022). They argue that splitting the demand between two modes is efficient when there is heterogeneity among users' preferences, and when the modes do not have great levels of scale economies. In our case:

- The heterogeneity of the users does not come from their subjective preferences, but from their walking distance towards the fixed line. Therefore, this condition is fulfilled.
- The fixed line does present scale economies that are partially lost because some users are shifted to ODRP. As argued above, this barely increases waiting times (in other words, the demand levels are large enough for the Mohring Effect to be almost exhausted). In the case of ODRP, it does present overall scale economies, but for users it presents a mixture between scale economies and diseconomies (Fielbaum et al., 2023). All together, this suggests that there are losses of scale economies when users are split, but they are limited.

Real-life bus frequencies: Finally, let us study a different scenario, namely comparing our approach to currently-offered real-life (rather than optimized) bus frequencies. We show here that our results get much better. That is, here the original scenario considers the bus frequencies per period as exogenous and equal to the actual bus frequencies set for each line by the public transportation authorities. The whole method is applied in the same way afterwards: capacities are calculated to fit the demand, and after

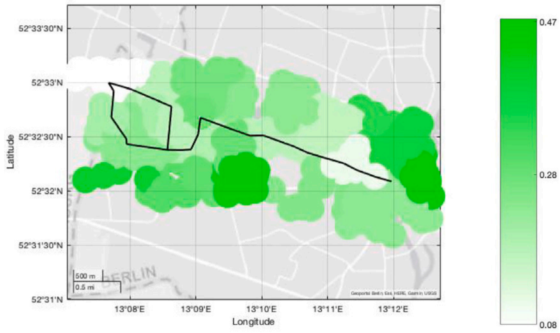


Fig. 7. Fraction of the users emerging from every zone that are served by ODRP, for the Line 237 in Berlin with 20 flexible vehicles. The greener the area, the more intense the use of ODRP. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

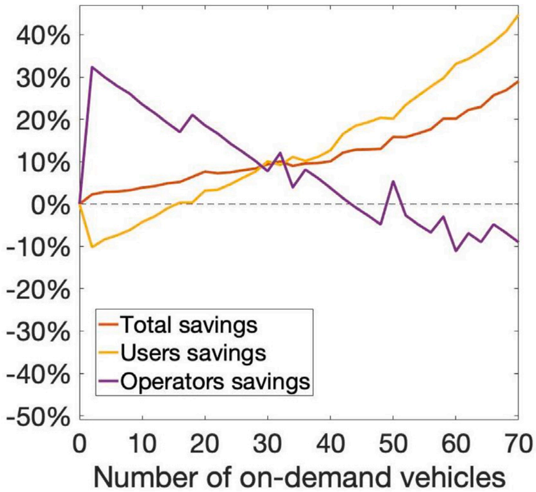


Fig. 8. Percentage of the costs saved thanks to utilizing the mixed system, compared to the original case with no flexible vehicles, considering real-life frequencies for Line 237 in Berlin.

introducing ODRP the frequencies are reduced maintaining the vehicles' capacities. The savings reported below are compared to the total costs obtained with these frequencies.

It is worth noting that the comparison provided before is fairer than this one, as the total savings are calculated there assuming the same cost function utilized for optimizing the frequency. Therefore, the results here are meant to show the robustness of our approach. We show the results in Berlin, as the trends are very clear and the reasons behind them remain the same for any other scenario.

Results are depicted in Fig. 8, and the trends are very strong. Savings increase in a very significant way for both users and operators compared to the case with optimized frequencies, which leads to a substantial increase in total savings as well. The main reasons for that are twofold:

- First, as stated by Jansson (1980), real-life frequencies are usually sub-optimal because "more buses should be run and buses should be much smaller". In this particular case, buses need capacity of at least 102 to be able to serve everybody with real-life frequencies, whereas the capacity becomes 29 when using optimal frequencies. This means that waiting times are much larger, so that users are greatly benefited by the introduction of ODRP.
- The utilization of large vehicles also lead to greater idle capacity during the off-peak periods (Jara-Díaz et al., 2017). Therefore, when we calculate the updated frequencies after introducing ODRP, the room for reducing them without violating the bus capacity is larger than in the optimal-frequencies case, which leads to a reduction in the operators' costs.

4. Limitations of the model and implementation challenges

In this section, we discuss the limitations of our model and simulations. We first focus on the model itself, recalling what were the main assumptions we made and what impact they have on the results (Section 4.1). Then we focus on what challenges would be faced to implement the proposed system in practice, and how this would relate to the rest of the public transportation system.

4.1. Modeling limitations

In this section, we discuss the four assumptions listed when describing the model at the beginning of Section 2.

(A1) **OD Matrix fixed and deterministic:** The operational characteristics of our system, i.e. its fleets and frequencies, are tailored to the specific demand of the day we are modeling. In practice, (i) public transportation systems do not know the actual passenger demand in advance, and therefore must offer services based on forecasts that are usually estimated with historical data, and (ii) an improved public transportation system could attract more demand. We now argue that this assumption does not affect our overall conclusions.

Let us start with the problem of known and deterministic demand. First, this affects both the original scenario (only buses) and ours (buses and ODRP). Therefore, if (for instance) more vehicles would be required to face random demand variations, this would increase the costs of both scenarios. Moreover, by design ODRP is better suited to face random demand variations, especially variability in the spatial location of demand, precisely because it is on-demand, i.e., the vehicles have more flexibility to adjust. As such, dropping this assumption would likely lead to an increase in costs for both scenarios, but we expect that the comparison would favor the proposed model even more than in the current setting. Second, it is worth remarking that this is a usual assumption in models that decide frequencies and/or networks in the context of traditional public transportation (see the survey by Durán-Micco and Vansteenwegen (2022) for a more detailed discussion). This reinforces the arguments given above: our findings are valid from a strategic perspective, but as most public transportation design models, a real-life implementation would require designing complementary methods to forecast the demand.

We now focus on the problem of non-reactive demand. As the proposed system offers a better quality of service than the original one, more users would be attracted to public transportation as a whole. Moreover, most of the new users would likely be situated far from the fixed line, as those are the ones most benefited from the presence of ODRP. This means that the spatial distribution of the demand might change: some intrazonal trips are regarded as feasible in our model, as long as the origin and destination are at walking distance from the fixed line. In reality, some users might avoid taking a bus for an intrazonal trip (e.g., if the walking time is too long compared to the total duration of the trip, so that a private mode could be much more attractive), but accept ODRP for the same purpose (Zuniga-Garcia et al., 2022). While this means that our results should be updated to account for new users,[11] doing so would only improve their travel situation: first, because attracting more users to public transportation is usually a desirable goal, and second, because ODRP is most-likely characterized by overall scale economies (Fielbaum et al., 2023).

Two nuances emerge: First, ODRP's quality of service might become worse, under certain conditions, when its ridership increases[12] especially if the spatial distribution of the demand changes. Second, the emergence of some users that would accept traveling via ODRP but not in the fixed line (the "intrazonal users" discussed above) imply that some passengers, if not assigned to an ODRP vehicle, would leave the public transportation system and find a private way to travel. Both aspects would mitigate the increase in ridership but do not affect the overall argument.

Finally, we remark that using the current public transport demand to represent the demand of a system that combines on-demand mobility with fixed lines, as well as assuming a non-reactive demand even when the combination changes (i.e., more ODRP or more fixed lines), are commonly observed in the literature (Leich and Bischoff, 2019; Maheo et al., 2019; Calabrò et al., 2023; Ng et al., 2024a; Edirimanna et al., 2024; Fielbaum and Alonso-Mora, 2024). A more sophisticated (endogenous) demand model would make our results more accurate and is regarded as a relevant direction for future research.

(A2) **Congestion is disregarded:** In our model we do not account for traffic congestion, which could be included by either exogenous or endogenous factors.

Exogenous factors refer to the fact that the conditions on the streets vary during the day (for instance, if travel times are longer during peak periods because of car traffic). This problem would affect exactly the same way both scenarios (the original and our proposed system), and thus our conclusions are not affected.

Endogenous factors refer to the extra congestion that would be caused by the ODRP vehicles in the system. They do increase the VKT and Vehicles-Hours-Traveled (VHT). This is relevant not only to achieve a more accurate model, but also because it may raise environmental concerns. E.g. in the Berlin case we study in-depth in Section 3 (i.e., line 237 with 20 vehicles), VHT increases from 442 to 655. However, this is a mild increase when compared to the overall car-dominated traffic volume,[13] which would be likely overweighed by the changes in mode choice. Similarly, Line 501 in Santiago covers approximately 11 km per direction, so 22 km in total, where the streets have at least two lanes in every part of its route. In the largest ODRP fleet scenario (80 flexible vehicles), without considering that some buses are removed, and in a pessimistic case where all of them are concentrated in the same streets used by the buses, this would add 1.8 veh/km-lane, which is a very mild number compared to the more than 100 veh/km-lane required to achieve a jam density[14] (Knoop and Daamen, 2017).

[11] We remark that no method trying to predict the demand changes can be very accurate, because there is no real-life implementation of our system to have a data-oriented estimation. Therefore, we opt for a clear and crisp method, namely keeping the same OD matrix, where we know the direction of the bias — disregarding potential new trips.

[12] When more users share the same vehicle, each of them faces a longer detour. This is defined as the *Extra Detour effect* by Fielbaum et al. (2023). Similarly, short trips would make the ODRP vehicles to stop more often.

[13] Berlin has less than 200 bus lines, so a rough approximation is that this system would add less than 5000 thousands vehicles to the streets. On the other hand, cars account for 45% of the trips in Berlin, thus almost a million cars. The impact would be even milder in most other cities, as Berlin has a relatively low private transportation share.

In other words, if including ODRP is able to attract car drivers as discussed in the previous bullet point, the congestion and emissions could actually decrease (in fact, using empirical data, Tirachini et al. (2020) shows that the actual effect of ridepooling on increasing or reducing vehicle kilometers traveled critically depends on the occupancy rate of ridepooling trips and on ridepooling's ability to attract car drivers).

Finally, we remark that assuming fixed traveling times is frequent among the literature dealing with ODRP and its integration into public transport (Maheo et al., 2019; Simonetto et al., 2019; Pinto et al., 2020; Calabrò et al., 2023; Ng et al., 2024b). Accounting for the endogenous congestion is methodologically challenging (Salazar et al., 2019), and is regarded as a relevant direction for further research.

(A3) **Single value of time is representative of the demand:** The critical role played by this assumption is that we know in advance how each user perceives the quality of service in the fixed line. This enables having Eq. (12) as a simple rule to prioritize the users that face the worst quality of service. If the true values of time of each user was known, it is straightforward to adapt Eq. (12) so that it still reflects the fixed line's quality of service. However, this could raise equity issues (e.g., is it fair to prioritize someone just because her value of time is larger, even if her objective travel time is smaller?). Moreover, knowing the value of time for each user is typically impossible. This is why we regard using average values of time for everyone as reasonable.

Dropping this assumption might imply that for some users our rule might no longer present incentive compatibility (Proposition 1 may not be valid for them). To be specific, if a user enjoys walking but hates waiting, then our method might be too optimistic regarding how she would perceive the offered ODRP alternative. However, this can be easily solved by allowing users to reject a trip when it is offered (but not after being accepted, as discussed when analyzing users' behavior in Section 4.2) and the rest of the method would function almost the same.

(A4) **All users first request ODRP:** Some users might prefer to directly use the bus instead of requesting an ODRP first. This can happen for at least two different reasons. First, if the fixed line works well for a user, she will know that she will not be prioritized when assigning the flexible vehicles, so going directly to the bus stop is more convenient. Second, some users (e.g. the elderly) might face a technological barrier that prevents them from using apps, or might prefer to continue using the buses as a matter of habit or personal preference. The former case is not a problem, as this behavior would actually improve our results (by deducting the extra $\delta/2$ waiting for those users). The latter issue is more problematic, as those users would be in a slightly worse situation than when the ODRP service does not exist. However, and as described in Section 3, the increase on average waiting times is small (in our simulations, less than a minute), so this is far from being a major issue.

4.2. Implications for practice

The promising results of the system proposed here suggest that this is an appealing way to introduce ODRP into public transportation networks. However, there is usually friction between these models and a potential real-life implementation, as many aspects of reality are more complex than what we have discussed so far. Let us discuss now the most important challenges and changes that would need to be considered for an eventual implementation of this idea:

- Users' behavior: A frequent problem observed in shared on-demand mobility is the direct effects of users' behavior on their co-travelers. Concretely, if users cancel a trip, or show up late, the system is affected as a whole (Kucharski et al., 2020; Yao and Bekhor, 2024). Private platforms usually prevent this by charging the users even if the trip does not take place, so similar measures would need to be considered if a user assigned to ODRP, after accepting the trip, cancels or is late; with or without these penalties, the quantitative findings of the paper would change when accounting for potential cancellations and late users. In order to ease the visual connection between matched ODRP vehicles and passengers, it might be beneficial to limit the number of nodes in the network that can be used as pickup or dropoff points, and provide some basic infrastructure there (e.g. a sign). Finally, the transition towards this type of system could be difficult for the users, as they are accustomed to knowing in advance which type of vehicle they will board: therefore, the gains in quality of service should be made very explicit for the users to be willing to explore this new alternative.
- As discussed in the introduction, other ways to integrate ODRP have been proposed in the past, such as using ODRP to face the last-mile problem, or to fully replace traditional buses in low-demand areas. We envision that at a whole-city level, these different strategies can be combined depending on which of them suits each area better. In particular, our results show that in urban areas currently served by a fixed line, i.e., where many users are traveling longitudinally, the system proposed here can be an efficient solution. On the other hand, low-demand areas where origins and destinations are dispersed might still be better served by ODRP only, while in suburban zones with a rail station, it might be better to use ODRP as a feeder.

[14] Note that vehicles per lane are not enough to describe congestion, as these shared vehicles stop more often, and large vehicles or buses can create sight problems. Two formal ways to deal with an increased number of stops are to assume larger equivalence factors for shared-rides vehicles in mixed-traffic operations, because their stops to pick up and drop off passengers may delay all other vehicles (as done in Section 3.5 of Tirachini et al. (2014)), or to directly add to cars a fraction of the time buses are stopped at bus stops (as done by Basso and Silva (2014)). However, as mild as our numbers are (1.8 veh/km-lane), we do not model any direct impact (positive or negative) on congestion potentially induced by our system, and leave this analysis as a line of further research.

- Let us discuss what would happen when a user r is currently taking the fixed line to connect with another public transportation line. This means that our model, and specifically our OD matrix, would only capture one leg of r's trip. From the users' perspective, r would need to follow exactly the same sequence of public transport lines, but in some of them she might be assigned to a flexible vehicle instead of a bus.
  In those cases, if r is assigned to an ODRP vehicle, she would be dropped off in the correct place to board the second line. If the latter is also a bus line and is implementing this same strategy, this could imply a transfer between two ODRP vehicles. This is not a problem, but it does create opportunities for an improved efficiency, if the dropoff of the first vehicle can be coordinated with the pickup of the second one. However, including transfers in on-demand mobility is a very complex mathematical problem (Coltin and Veloso, 2014), so it is regarded as a direction for future research.
- Finally, and as discussed above, our system has better chances of being implemented when automated vehicles are available. While total costs are still significantly diminished in the human-driven scenario, the additional monetary costs might make the system unfeasible for now.

Besides practical challenges, our results suggest strategic insights that can be utilized when designing other types of integration between ODRP and fixed routes. Our model performs well because we are able to identify which users are currently receiving a low quality of service in public transportation (mostly those with large walking times), to prioritize them in the design of ODRP services. This is a promising idea in areas around a fixed public transportation line, as shown by the diverse case applications in this paper. We note that the popular idea of ODRP as a first-mile connector follows the same rationale: if with a fixed line some users would receive a low quality of service because of excessive access/egress times, ODRP could be useful as an access/egress mode. A new insight of this study is that this rationale can also be applied in high-demand areas. As such, besides the idea proposed here of complementing one traditional fixed line, ODRP could play other roles in urban environments following the same principle, e.g., prioritizing users that need to walk too much, face long detours, and possibly also needing several transfers. This is a very relevant direction for future research.

5. Conclusions

In this paper, we have proposed a novel way to integrate shared on-demand mobility into a public transportation system, which leverages the current fixed-lines structure of transit networks and complements each line with a fleet of small vehicles following flexible routes. We show how to adapt a state-of-the-art assignment algorithm for ODRP, so that it is more likely to serve the users whose fixed-line alternative is worse due to long walking times, and ensure that the system's decisions are aligned with users' interests. The fixed-line fleet can be reduced as now buses have fewer riders, compensating (at least partially) the increased operators' costs.

We test the efficiency of this system by simulating the daily operation of four different lines from Santiago, Chile, and Berlin, Germany, using data that comes from real-life operations. Our results show that users' average costs can be greatly reduced, namely by serving passengers located far from the bus stops (preventing long walking times), and by fully replacing the fixed-line during the lowest-demand periods (preventing long waiting times). Further, if vehicles are automated, a low number of small vehicles can also reduce operators costs (i.e., this system can obtain a Pareto improvement), by preventing the utilization of large vehicles with idle capacity during the low-demand periods. The optimal ODRP fleet reduces total costs in 13%–39%, depending on the line, compared to using the traditional fixed-line only.

From an operational perspective, our results suggest that given a zone to be covered by public transport, some passengers are better served through a traditional fixed route, when their origins and destinations are close to a large street, while others are better off in ODRP. In densely populated areas, it is possible to have a demand high enough so that both sub-systems can operate efficiently — offering low waiting times in the fixed-line, and matching the ODRP passengers in groups with limited detours. This reinforces that ODRP should not be limited to operating in low-demand contexts

There are many research lines to extend our results. First, in this paper we have assumed that ODRP collaborates with the fixed line: How much would be lost if they compete, i.e., if ODRP decides whom to serve trying to maximize its profit instead of prioritizing the users worsely served by the fixed line? Second, it is crucial to complement this supply-based analysis with demand models, in order to determine if the increased quality of service would be enough to attract some car users to this mixed system. Third, dynamic assignment methods that account for the changes in the exogenous levels of traffic and congestion would make our results more robust. Finally, in our model the flexible vehicles are constrained to operate in the area served by a public transportation line; despite the virtues discussed in the paper, there might be several other alternatives to complement the ideas discussed here. In other words, as we have shown that this mixed service outperforms the original fixed-line and is also better than using only ODRP, we conclude that the main question to be investigated is not how to choose between a pure traditional system or a pure on-demand one, but how to combine the best of both.

CRediT authorship contribution statement

Andres Fielbaum: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alejandro Tirachini:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Formal analysis, Conceptualization. **Javier Alonso-Mora:** Writing – original draft, Supervision, Resources, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Acknowledgments

The authors thank the Directory of Metropolitan Public transportation (DTPM) for providing the Santiago data, and Martin Meyer (TU Berlin) for providing the Berlin data for this research. We are also indebted to Felipe Vera (U. de Chile) and Maximilian Kronmueller (TU Delft) for their help to build the graphs of Santiago and Berlin. We thank the anonymous comments by four reviewers that significantly helped improve the contents of this paper.

Appendix A

Details of the ODRP assignment method

Here we explain how to compute the set of feasible trips $\mathcal{T}$, which is done following the method by Fielbaum et al. (2021). Recall that this set, as well as the subsequent assignments, are computed in a receding horizon fashion. This means that every Δt we accumulate the users that emerge during that lapse of time and assign them at once, together with users that have been previously assigned but not yet picked up, so that we allow for reassignments. We use the same rule as Fielbaum et al. (2021), where users that have been told to walk towards a pick-up point different than their origin will not be reassigned.

We denote by R_t the set of requests to be assigned at time t, and $\mathcal{V}$ the set of vehicles. When computing $\mathcal{T}$, each $v \in \mathcal{V}$ is characterized by its position $Pos(v,t)$, its planned route $route(v,t)$, and the requests currently assigned to it $\zeta(v,t)$. A trip T, that assigns the requests $req(T)$ to the vehicle $veh(T)$, is feasible if there exists pick-up and drop-off nodes for each request in $req(T)$, i.e., $\{pu_r, do_r : r \in req(T)\}$, and an updated route π for $veh(T)$, such that π visits all such pick-up and drop-off nodes and the ones previously assigned to the requests in $\zeta(v,t)$, and fulfills the feasibility constraints. We assume that the vehicle spends a fixed time δ_{stop} per stop due to breaking and accelerating.

In order to define the feasibility constraints, let us denote $Wait(r,\pi)$ and $IV(r,\pi)$ the waiting time and in-vehicle time, respectively, that would be faced by request r when the vehicle follows the route π. Moreover, let us define $cap(r,T,\pi)$ the maximum number of users that would share the vehicle at the same time. Then, the following restrictions have to be fulfilled:

$$\begin{aligned}
t_W(o_r, pu_r) &\leq \Omega_{Walk} && \forall r \in req(T)\\
t_W(d_r, do_r) &\leq \Omega_{Walk} && \forall r \in req(T)\\
Wait(r,\pi) &\leq \Omega_{Wait} && \forall r \in req(T)\\
IV(r,\pi) + Wait(r) + Walk(r) - t_V(o_r, d_r) &\leq \Omega_{Delay} && \forall r \in req(T)\\
cap(r,T,\pi) &\leq \kappa
\end{aligned}$$

where the parameters $\Omega_{Walk}, \Omega Wait$, and Ω_{delay} represent the maximum admissible walking, waiting, and total delay. The first four restrictions represent the hard time windows, while the final one implies that the capacity of the vehicle can never be exceeded. We assume $\Omega_{Walk} = \Omega_{Delay} = 20$ min. Regarding waiting time, we mimic what happens in public transport: if the user belongs to the period p, then its maximum waiting time is $1/F(p)$.

These time windows not only ensure a reasonable quality of service, but also restrict the number of feasible trips. Nevertheless, finding all of them is still challenging from a combinatorial point of view, which is why we leverage the following property first identified (and proved) by Alonso-Mora et al. (2017): a necessary condition for the trip $(req(T), veh(T))$ to be feasible, is that $\forall H \subset req(T)$, the trip $(H, veh(T))$ is feasible as well. Using this property, we identify the feasible trips incrementally, as we now explain: For each $v \in \mathcal{V}$, we first search for all the individual requests (*groups* of $k = 1$ requests) that could be feasibly be served by v. When we know all the groups of size k that can be transported by v, we study the groups of size $k+1$, but considering only those such that the subgroups of size k are feasible.

To determine whether a given group can be feasibly served by v, we should try all the possible routes, and every possible pick-up and drop-off points for each request. Instead, we use an insertion heuristic for the routes, and we use a local search to determine candidates for the pick-up and drop-off locations. Both heuristics are explained in detail by Fielbaum et al. (2021).

Putting everything together, we are able to determine all the feasible trips at time t. Once this is done, the trips that are actually being assigned are determine through the ILP explained in Section 2.

Proof of Proposition 1. Suppose there is $r \in req(T)$ such that $UC_T(r) > UC_{Fixed}(r)$. Consider the trip T', with $veh(T') = veh(T)$ and $req(T') = req(T) \setminus \{r\}$. Let us consider (x', z') an alternative solution, where $x'_T = 0, x'_{T'} = 1$, $z'_r = 1$, and everything else is unmodified. We will show that (x', z') is feasible and yields a lower value of the objective function, which is a contradiction as the original solution was supposed to be optimal. Note that, in plain words, the proposed modification of the solution just implies taking r out of the itinerary of $veh(T)$.

The new solution is feasible: This is straightforward to verify. Eq. (10) is unmodified for every request not in $req(T)$. Requests in $req(T)$ other than r belong to $req(T')$, so that Eq. (10) is valid for them because $x'_{T'} = 1$ and $x'_T = 0$. As now r is unserved, the second addend in Eq. (10) becomes zero but $z'_r = 1$.

We now analyze the vehicles and Eq. (11). Note that this equation is unmodified for all the vehicles other than $v = veh(T)$. In the case of v, as $x'_T = 0$ and $x'_{T'} = 1$, the result of the sum does not change.

The new solution yields a lower value for the objective function: The objective function is modified by

$$\Delta C = C(x', z') - C(x, z) = \psi(r) + UC(T') - UC(T) \tag{14}$$

We need to show that $\Delta C < 0$. We do this by splitting ΔC into two terms, $\Delta C = \Delta_1 C + \Delta_2 C$, with

$$\Delta_1 C = \psi(r) - UC_T(r), \tag{15}$$

and

$$\Delta_2 C = \sum_{s \in req(T')} UC_{T'}(s) - UC_T(s) + \sum_{s \in Pax(v)} \Delta UC_{T'}(s) - \Delta UC_T(s) \tag{16}$$

We know that $\Delta_1 C < 0$, as this is exactly the hypothesis we assume to obtain a contradiction at the very beginning of this proof. On the other hand, $\Delta_2 C \leq 0$ thanks to the triangular inequality, which ensures that if we are to serve a strict subset of the passengers, we can do it in a less costly way. This can be seen by recalling that the vehicle's route is optimized to minimize $UC(T')$ when serving T'; note that the route followed by v to serve T is also feasible to serve T', meaning that the associated cost, which is exactly $\sum_{s \in req(T')} UC_T(s) + \sum_{s \in Pax(v)} \Delta UC_T(s)$, is an upper bound for the cost of the optimal route, which is $\sum_{s \in req(T')} UC_{T'}(s) + \sum_{s \in Pax(v)} \Delta UC_{T'}(s)$. $\square$

Details of the computation of the exact origins and destinations

Recall that for all the scenarios we study, the datasets reveals where users board and alight the buses, but not the exact origins and destinations. In other words, if we assume (w.l.o.g.) that they travel in direction S, we know the stops S_i and S_j where the trip begins and finish, respectively. This information is not enough to simulate the ODRP operation, so for each request r we computed an exact origin $o_r \in V$ and an exact destination $d_r \in V$.

Let us explain how do we compute o_r for a request r that boards the bus in S_i, as the method for d_r is exactly the same. We first identify the set of nodes $V_i \subset V$ fulfilling $V_i = \{v \in V : t_W(v, S_i) \leq t_W(v, S_j) \forall j = 1, \dots, n\}$. In other words, the set V_i contains all the nodes that would have chosen the stop S_i, i.e., the nodes we consider as candidates to be o_r. For each $v \in V_i$, define

$$\rho_v = \frac{1}{\Theta + t_W(v, S_i)}$$

After normalization, ρ_v defines the probability that $o_r = v$. Therefore, those nodes that are situated closer to the bus station will have a greater probability of being selected as the origin. The parameter Θ is a constant number, which is needed because otherwise we would obtain $\rho_{S_i} = +\infty$. We use $\Theta = 2$ min, i.e., one tenth of the maximum admitted walking time (20 min, as explained in Section 3).

Putting everything together, if r would board the bus in S_i, the probability that the exact origin of r is $v \in V_i$ is given by:

$$P(o_r = v) = \frac{\rho_v}{\sum_{u \in V_i} \rho_u}$$

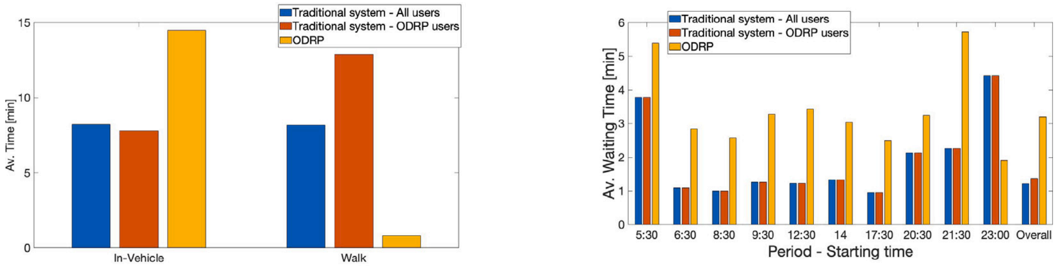
It is worth mentioning that we regard this formula, and the chosen value of Θ, as realistic. As shown in Fig. 4, the average walking time towards the fixed line that is achieved by our system is approximately 8 min. This is less than the average times reported in the US (Besser and Dannenberg, 2005), and similar to the times reported in Munich other than in the CBD (Sarker et al., 2020) — and we do not study any line internal to a CBD in our applications. It is worth noting that these average times depend strongly on the city and system (Van Soest et al., 2020).

Numerical value of the parameters

In-vehicle time savings are taken from Tirachini and Antoniou (2020). Values of waiting and access time are weighted by 2 and 2.5, respectively, as done by Fielbaum et al. (2021). As discussed in the main text, the parameters related to operators' costs are also derived from Tirachini and Antoniou (2020), who estimate the different components of the cost function (such as maintenance, energy, and capital) for $K = 5, 8, 50, 90, 140$, in Santiago, Chile, and Munich, Germany, for human-driven and automated electric vehicles. We assume the values in Munich to be valid for Berlin as well. We classify each of such components to be either dependent on the number of driving hours (hence contributing to c_{KO} and c_{BO}) or not (contributing to c_{KC} and c_{BC}). Numbers are normalized to account for the number of driving hours per day. After this procedure, we have total fixed and usage-dependent costs for each $K = 5, 8, 50, 90, 140$, so that a linear regression leads to the parameters shown in Table 6. The times needed to break, accelerate, and open doors, are computed following Roess et al. (2004) and Tirachini (2014), which consider the vehicles' type and velocity, while t_0 is taken from Fielbaum et al. (2018).

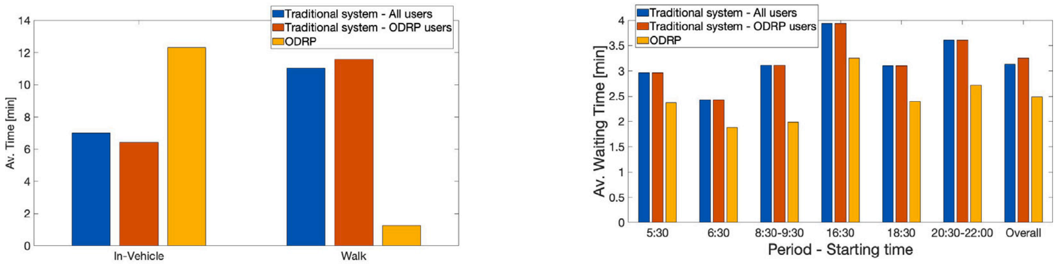
Table 6
Numerical value of the parameters.

Parameter	Value in Santiago	Value in Berlin
Users' costs parameters		
α_{Walk}	7.3 [$US \h]	13 [$US \h]
α_{Wait}	5.8 [$US \h]	10.4 [$US \h]
α_{IV}	2.9 [$US \h]	5.2 [$US \h]
Operators' cost parameters: General		
c_{BO}	1.13 [$US \h]	1.13 [$US \h]
c_{KO}	0.043 [$US \h-pax]	0.074 [$US \h-pax]
Automated vehicles		
c_{BC}	30.28 [$US]	24.6 [$US]
c_{KC}	1.1 [$US \pax]	2.1 [$US \pax]
Human-driven vehicles		
c_{BC}	78.9 [$US]	157.8 [$US]
c_{KC}	1.2 [$US \pax]	2.4 [$US \pax]
Other parameters		
Time to break, accelerate, open doors t_{BOA} (large buses)	13 [s]	13 [s]
Time to break and accelerate ODRP	5 [s]	5 [s]
t_0	5 [s]	5 [s]



(a) (b)

Fig. 9. Comparison of service times in ODRP and in the system without ODRP for the Line 501 in Santiago with 6 vehicles.



(a) (b)

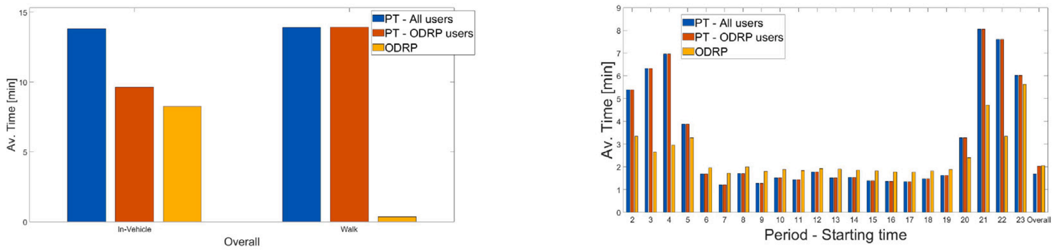
Fig. 10. Comparison of service times in ODRP and in the system without ODRP for the Line 208C in Santiago with 6 vehicles.

Additional figures in the optimal-frequency scenario

Figs. 9–11 are equivalent to Fig. 5 for the three other scenarios. As explained above, the same conclusions hold, namely that the users that are mostly benefited by this integrated system are the ones that would face a long walking time if using the traditional line, or the ones traveling in low-demand periods. It is noteworthy that in-vehicle time does not play a major role here, and that in the case of Line 208C, all the ODRP users are benefited through waiting time, which happens because this is a low-demand (hence low-frequency) line.

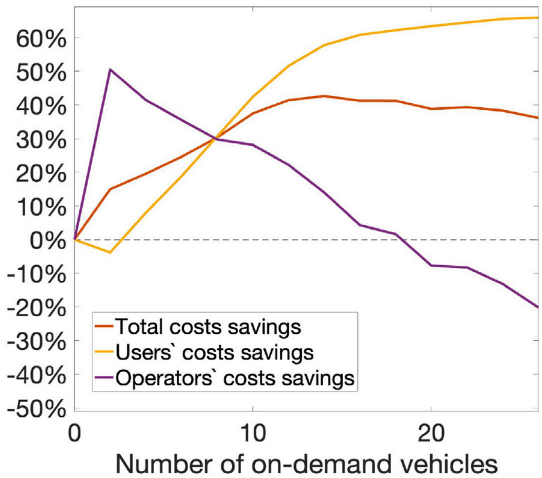
Additional figure for a reduced sample

See Fig. 12.



(a) (b)

Fig. 11. Comparison of service times in ODRP and in the system without ODRP for the Line M45 in Berlin with 48 vehicles.



0.48

Fig. 12. Percentage of the costs saved thanks to utilizing the mixed system with automated vehicles in Line 237 in Berlin with half of the demand, compared to the original case with no flexible vehicles.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.tra.2024.104289.

References

Agarwal, Saharsh, Mani, Deepa, Telang, Rahul, 2023. The impact of ride-hailing services on congestion: Evidence from Indian cities. Manuf. Serv. Oper. Manage..

Alonso-González, María J, van Oort, Niels, Cats, Oded, Hoogendoorn-Lanser, Sascha, Hoogendoorn, Serge, 2020. Value of time and reliability for urban pooled on-demand services. Transp. Res. C 115, 102621.

Alonso-Mora, Javier, Samaranayake, Samitha, Wallar, Alex, Frazzoli, Emilio, Rus, Daniela, 2017. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. Proc. Natl. Acad. Sci. 114 (3), 462–467. http://dx.doi.org/10.1073/pnas.1611675114.

Badia, Hugo, Jenelius, Erik, 2020. Feeder transit services in different development stages of automated buses: comparing fixed routes versus door-to-door trips. Transp. Res. Procedia 47, 521–528.

Bahamonde-Birke, Francisco J., 2022. Should competition between regulated public transport and autonomous ride-sharing providers be allowed? An outlook into a possible transport paradox. J. Transp. Econ. Policy (JTEP) 56 (1), 56–78.

Basso, Leonardo J., Silva, Hugo E., 2014. Efficiency and substitutability of transit subsidies and other urban transport policies. Am. Econ. J.: Econ. Policy 6 (4), 1–33.

Basu, Rounaq, Araldo, Andrea, Akkinepally, Arun Prakash, Nahmias Biran, Bat Hen, Basak, Kalaki, Seshadri, Ravi, Deshmukh, Neeraj, Kumar, Nishant, Azevedo, Carlos Lima, Ben-Akiva, Moshe, 2018. Automated mobility-on-demand vs. mass transit: a multi-modal activity-driven agent-based simulation approach. Transp. Res. Rec. 2672 (8), 608–618.

Baum-Snow, Nathaniel, 2007. Did highways cause suburbanization? Q. J. Econ. 122 (2), 775–805.

Ben-Akiva, Moshe, Morikawa, Takayuki, 2002. Comparing ridership attraction of rail and bus. Transp. Policy 9 (2), 107–116.

Besser, Lilah M., Dannenberg, Andrew L., 2005. Walking to public transit: steps to help meet physical activity recommendations. Am. J. Prev. Med. 29 (4), 273–280.

Börjesson, Maria, Eliasson, Jonas, Rubensson, Isak, 2020. Distributional effects of public transport subsidies. J. Transp. Geogr. 84, 102674.

Borndörfer, Ralf, Grötschel, Martin, Pfetsch, Marc E., 2007. A column-generation approach to line planning in public transport. Transp. Sci. 41 (1), 123–132.

Bruzzone, Francesco, Scorrano, Mariangela, Nocera, Silvio, 2021. The combination of e-bike-sharing and demand-responsive transport systems in rural areas: A case study of Velenje. Res. Transp. Bus. Manage. 40, 100570.

Bürstlein, Johanna, López, David, Farooq, Bilal, 2021. Exploring first-mile on-demand transit solutions for North American suburbia: A case study of Markham, Canada. Transp Res. A 153, 261–283.
Calabrò, Giovanni, Araldo, Andrea, Oh, Simon, Seshadri, Ravi, Inturri, Giuseppe, Ben-Akiva, Moshe, 2023. Adaptive transit design: Optimizing fixed and demand responsive multi-modal transportation via continuous approximation. Transp. Res. A 171, 103643.
Carrese, Filippo, Sportiello, Simone, Zhaksylykov, Tolegen, Colombaroni, Chiara, Carrese, Stefano, Papaveri, Muzio, Patella, Sergio Maria, 2023. The integration of shared autonomous vehicles in public transportation services: A systematic review. Sustainability 15 (17), 13023.
Chen, Shukai, Wang, Hua, Meng, Qiang, 2020. Solving the first-mile ridesharing problem using autonomous vehicles. Comput.-Aided Civ. Infrastruct. Eng. 35 (1), 45–60.
Choi, Yunkyung, Guhathakurta, Subhrajit, Pande, Anurag, 2022. An empirical Bayes approach to quantifying the impact of transportation network companies (TNCs) operations on travel demand. Transp. Res. A 161, 269–283.
Coltin, Brian, Veloso, Manuela, 2014. Ridesharing with passenger transfers. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, pp. 3278–3283.
Daganzo, Carlos F., 1984. Checkpoint dial-a-ride systems. Transp. Res. B 18 (4–5), 315–327.
Daniels, Rhonda, Mulley, Corinne, 2013. Explaining walking distance to public transport: The dominance of public transport supply. J. Transp. Land Use 6 (2), 5–20.
Diao, Mi, Kong, Hui, Zhao, Jinhua, 2021. Impacts of transportation network companies on urban mobility. Nat. Sustain. 4 (6), 494–500.
Durán-Hormazábal, Elsa, Tirachini, Alejandro, 2016. Estimation of travel time variability for cars, buses, metro and door-to-door public transport trips in Santiago, Chile. Res. Transp. Econ. 59, 26–39.
Durán-Micco, Javier, Vansteenwegen, Pieter, 2022. A survey on the transit network design and frequency setting problem. Public Transp. 14 (1), 155–190.
Edirimanna, Danushka, Hu, Hins, Samaranayake, Samitha, 2024. Integrating on-demand ride-sharing with mass transit at-scale. arXiv preprint arXiv:2404.07691.
Ferretto, Laura, Bruzzone, Francesco, Nocera, Silvio, 2021. Pathways to active mobility planning. Res. Transp. Econ. 86, 101027.
Fielbaum, Andrés, 2020. Strategic public transport design using autonomous vehicles and other new technologies. Int. J. Intell. Transp. Syst. Res. 18 (2), 183–191.
Fielbaum, Andrés, 2021. Optimizing a vehicle's route in an on-demand ridesharing system in which users might walk. J. Intell. Transp. Syst. 1–20.
Fielbaum, Andres, Alonso-Mora, Javier, 2020. Unreliability in ridesharing systems: Measuring changes in users' times due to new requests. Transp. Res. C 121, 102831.
Fielbaum, Andres, Alonso-Mora, Javier, 2024. Design of mixed fixed-flexible bus public transport networks by tracking the paths of on-demand vehicles. Transp. Res. C 104580.
Fielbaum, Andres, Bai, Xiaoshan, Alonso-Mora, Javier, 2021. On-demand ridesharing with optimized pick-up and drop-off walking locations. Transp. Res. C 126, 103061.
Fielbaum, Andres, Jara-Diaz, Sergio, Alonso-Mora, Javier, 2024. Beyond the last mile: Different spatial strategies to integrate on-demand services into public transport in a simplified city. Public Transp. http://dx.doi.org/10.1007/s12469-023-00348-1.
Fielbaum, Andres, Jara-Díaz, Sergio, Gschwender, Antonio, 2018. Transit line structures in a general parametric city: The role of heuristics. Transp. Sci. 52 (5), 1092–1105.
Fielbaum, Andrés, Pudāne, Baiba, 2024. Are shared automated vehicles good for public-or private-transport-oriented cities (or neither)? Transp. Res. D 136, 104373.
Fielbaum, Andrés, Tirachini, Alejandro, Alonso-Mora, Javier, 2023. Economies and diseconomies of scale in on-demand ridepooling systems. Econ. Transp. 34, 100313.
Fielding, Gordon J., 1975. Demand-responsive transit: an overview of integrated systems. Transportation.
Garcia-Martinez, Andres, Cascajo, Rocio, Jara-Diaz, Sergio R, Chowdhury, Subeh, Monzon, Andres, 2018. Transfer penalties in multimodal public transport networks. Transp. Res. A 114, 52–66.
Glaister, Stephen, 2018. The allocation of urban public transport subsidy. In: Privatisation and the Welfare State. Routledge, pp. 177–200.
Gössling, Stefan, Schröder, Marcel, Späth, Philipp, Freytag, Tim, 2016. Urban space distribution and sustainable transport. Transp. Rev. 36 (5), 659–679.
Gronau, Reuben, 2000. Optimum diversity in the public transport market. J. Transp. Econ. Policy 21–41.
Guenther, Karl W., 1976. Demand-responsive transportation in Ann Arbor: Operation. Transp. Res. Rec. 608, 20–25.
Gurumurthy, Krishna Murthy, Kockelman, Kara M., 2022. Dynamic ride-sharing impacts of greater trip demand and aggregation at stops in shared autonomous vehicle systems. Transp. Res. A 160, 114–125.
Henao, Alejandro, Marshall, Wesley E., 2019. The impact of ride-hailing on vehicle miles traveled. Transportation 46 (6), 2173–2194.
Hörcher, Daniel, Graham, Daniel J., 2021. The Gini index of demand imbalances in public transport. Transportation 48 (5), 2521–2544.
Hörcher, Daniel, Graham, Daniel J., 2022. Multimodal substitutes in public transport: Efficient variety or wasteful competition? J. Transp. Econ. Policy (JTEP) 56 (1), 1–27.
Huang, Yantao, Kockelman, Kara M., Garikapati, Venu, 2022. Shared automated vehicle fleet operations for first-mile last-mile transit connections with dynamic pooling. Comput. Environ. Urban Syst. 92, 101730.
Jansson, Jan Owen, 1980. A simple bus line model for optimisation of service frequency and bus size. J. Transp. Econ. Policy 53–80.
Jara-Díaz, Sergio, 2007. Transport Economic Theory, vol. 140, Elsevier Oxford, UK.
Jara-Díaz, Sergio, Fielbaum, Andrés, Gschwender, Antonio, 2017. Optimal fleet size, frequencies and vehicle capacities considering peak and off-peak periods in public transport. Transp. Res. A 106, 65–74.
Jara-Diaz, Sergio, Fielbaum, Andres, Gschwender, Antonio, 2020. Strategies for transit fleet design considering peak and off-peak periods using the single-line model. Transp. Res. B 142, 1–18.
Jara-Díaz, Sergio, Gschwender, Antonio, 2003. Towards a general microeconomic model for the operation of public transport. Transp. Rev. 23 (4), 453–469.
Jara-Díaz, S., Gschwender, A., 2009. The effect of financial constraints on the optimal design of public transport services. Transportation 36 (1), 65–75. http://dx.doi.org/10.1007/s11116-008-9182-8.
Jara-Diaz, Sergio, Gschwender, Antonio, Castro, Juan Cristobal, Lepe, Matias, 2024. Distance traveled, transit design and optimal pricing. Transp. Res. A 179, 103928.
Jiao, Guipeng, Ramezani, Mohsen, 2022. Incentivizing shared rides in e-hailing markets: Dynamic discounting. Transp. Res. C 144, 103879.
Ke, Jintao, Yang, Hai, Li, Xinwei, Wang, Hai, Ye, Jieping, 2020. Pricing and equilibrium in on-demand ride-pooling markets. Transp. Res. B 139, 411–431.
Ke, Jintao, Zhu, Zheng, Yang, Hai, He, Qiaochu, 2021. Equilibrium analyses and operational designs of a coupled market with substitutive and complementary ride-sourcing services to public transits. Transp. Res. E 148, 102236.
Kenworthy, J., 2003. Transport energy use and greenhouse gases in urban passenger transport systems: A study of 84 global cities. In: International Third Conference of the Regional Government Network for Sustainable Development, Notre Dame University, Fremantle, Western Australia.
Kikuchi, Shinya, 1984. Scheduling of demand-responsive transit vehicles. J. Transp. Eng. 110 (6), 511–520.
Kim, Jae Hong, Li, Xiangyu, 2021. Building more housing near transit: A spatial analysis of residential densification dynamics. Transp. Policy 114, 15–24.
Kirchhoff, Peter, 1995. Public transit research and development in Germany. Transp. Res. A 29 (1), 1–7.
Knoop, Victor L., Daamen, Winnie, 2017. Automatic fitting procedure for the fundamental diagram. Transportmetrica B: Transp. Dyn. 5 (2), 129–144.
Kucharski, Rafal, Fielbaum, Andres, Alonso-Mora, Javier, Cats, Oded, 2020. If you are late, everyone is late: late passenger arrival and ride-pooling systems' performance. Transportmetrica A: Transp. Sci. 1–24.

Lehe, Lewis, Pandey, Ayush, 2023. Bus stop spacing with heterogeneous trip lengths and elastic demand. Available at SSRN 4135394.
Leich, Gregor, Bischoff, Joschka, 2019. Should autonomous shared taxis replace buses? A simulation study. Transp. Res. Procedia 41, 450–460.
Li, Xiugang, Quadrifoglio, Luca, 2010. Feeder transit services: Choosing between fixed and demand responsive policy. Transp. Res. C 18 (5), 770–780.
Lucas, Karen, 2012. Transport and social exclusion: Where are we now? Transp. Policy 20, 105–113.
Ma, Mingfei, Jin, Ying, 2024. Estimating time-saving benefits and mode shifts from improvements of sustainable transport modes in Cambridge, UK. Case Stud. Transp. Policy 15, 101127.
Ma, Tai-Yu, Rasulkhani, Saeid, Chow, Joseph YJ, Klein, Sylvain, 2019. A dynamic ridesharing dispatch and idle vehicle repositioning strategy with integrated transit transfers. Transp. Res. E 128, 417–442.
Maheo, Arthur, Kilby, Philip, Van Hentenryck, Pascal, 2019. Benders decomposition for the design of a hub and shuttle public transit system. Transp. Sci. 53 (1), 77–88.
Munizaga, Marcela A., Palma, Carolina, 2012. Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. Transp. Res. C 24, 9–18.
Ng, Max, Dandl, Florian, Mahmassani, Hani S, Bogenberger, Klaus, 2024a. Semi-on-demand hybrid transit route design with shared autonomous mobility services. arXiv preprint arXiv:2403.15804.
Ng, Max TM, Mahmassani, Hani S, Verbas, Ömer, Cokyasar, Taner, Engelhardt, Roman, 2024b. Redesigning large-scale multimodal transit networks with shared autonomous mobility services. Transp. Res. C 104575.
Oke, Jimi B, Akkinepally, Arun Prakash, Chen, Siyu, Xie, Yifei, Aboutaleb, Youssef M, Azevedo, Carlos Lima, Zegras, P Christopher, Ferreira, Joseph, Ben-Akiva, Moshe, 2020. Evaluating the systemic effects of automated mobility-on-demand services via large-scale agent-based simulation of auto-dependent prototype cities. Transp. Res. A 140, 98–126.
Papanikolaou, Anestis, Basbas, Socrates, 2021. Analytical models for comparing demand responsive transport with bus services in low demand interurban areas. Transp. Lett. 13 (4), 255–262.
Pinto, Helen KRF, Hyland, Michael F, Mahmassani, Hani S, Verbas, I Ömer, 2020. Joint design of multimodal transit networks and shared autonomous mobility fleets. Transp. Res. C 113, 2–20.
Poinsignon, François, Chen, Lei, Jiang, Sida, Gao, Kun, Badia, Hugo, Jenelius, Erik, 2022. Autonomous vehicle fleets for public transport: Scenarios and comparisons. Green Energy Intell. Transp. 100019.
Quadrifoglio, Luca, Dessouky, Maged M, Ordóñez, Fernando, 2008. A simulation study of demand responsive transit system design. Transp. Res. A 42 (4), 718–737.
Quadrifoglio, Luca, Li, Xiugang, 2009. A methodology to derive the critical demand density for designing and operating feeder transit services. Transp. Res. B 43 (10), 922–935.
Roess, Roger P., Prassas, Elena S., McShane, William R., 2004. Traffic Engineering. Pearson/Prentice Hall.
Ronald, Nicole, Thompson, Russell, Winter, Stephan, 2015. Simulating demand-responsive transportation: a review of agent-based approaches. Transp. Rev. 35 (4), 404–421.
Salazar, Mauro, Tsao, Matthew, Aguiar, Izabel, Schiffer, Maximilian, Pavone, Marco, 2019. A congestion-aware routing scheme for autonomous mobility-on-demand systems. In: 2019 18th European Control Conference. ECC, IEEE, pp. 3040–3046.
Sarker, Rumana Islam, Mailer, Markus, Sikder, Sujit Kumar, 2020. Walking to a public transport station: empirical evidence on willingness and acceptance in Munich, Germany. Smart Sustain. Built Environ. 9 (1), 38–53.
Seriani, Sebastian, Fernandez, Rodrigo, Luangboriboon, Nattanon, Fujiyama, Taku, et al., 2019. Exploring the effect of boarding and alighting ratio on passengers' behaviour at metro stations by laboratory experiments. J. Adv. Transp. 2019.
Shen, Yu, Zhang, Hongmou, Zhao, Jinhua, 2018. Integrating shared autonomous vehicle in public transportation system: A supply-side simulation of the first-mile service in Singapore. Transp. Res. A 113, 125–136.
Shi, Kunbo, Shao, Rui, De Vos, Jonas, Cheng, Long, Witlox, Frank, 2021. The influence of ride-hailing on travel frequency and mode choice. Transp. Res. D 101, 103125.
Simonetto, Andrea, Monteil, Julien, Gambella, Claudio, 2019. Real-time city-scale ridesharing via linear assignment problems. Transp. Res. C 101, 208–232.
Stiglic, Mitja, Agatz, Niels, Savelsbergh, Martin, Gradisar, Mirko, 2018. Enhancing urban mobility: Integrating ride-sharing and public transit. Comput. Oper. Res. 90, 12–21.
Teal, Roger F., 1993. Implications of technological developments for demand responsive transit. Transp. Res. Rec. (1390).
Thorhauge, Mikkel, Jensen, Anders Fjendbo, Rich, Jeppe, 2022. Effects of autonomous first-and last mile transport in the transport chain. Transp. Res. Interdiscipl. Perspect. 15, 100623.
Tirachini, Alejandro, 2014. The economics and engineering of bus stops: Spacing, design and congestion. Transp. Res. A 59, 37–57.
Tirachini, Alejandro, Antoniou, Constantinos, 2020. The economics of automated public transport: Effects on operator cost, travel time, fare and subsidy. Econ. Transp. 21, 100151.
Tirachini, Alejandro, Chaniotakis, Emmanouil, Abouelela, Mohamed, Antoniou, Constantinos, 2020. The sustainability of shared mobility: Can a platform for shared rides reduce motorized traffic in cities? Transp. Res. C 117, 102707.
Tirachini, Alejandro, Gomez-Lobo, Andres, 2019. Does ride-hailing increase or decrease vehicle kilometers traveled (VKT)? A simulation approach for Santiago de Chile. Int. J. Sustain. Transp. 1–18.
Tirachini, Alejandro, Hensher, David A., 2011. Bus congestion, optimal infrastructure investment and the choice of a fare collection system in dedicated bus corridors. Transp. Res. B 45 (5), 828–844.
Tirachini, Alejandro, Hensher, David A., Rose, John M., 2014. Multimodal pricing and optimal design of urban public transport: The interplay between traffic congestion and bus crowding. Transp. Res. B 61, 33–54.
Tirachini, Alejandro, del Río, Mariana, 2019. Ride-hailing in Santiago de Chile: Users' characterisation and effects on travel behaviour. Transp. Policy 82, 46–57.
Van Soest, Dennis, Tight, Miles R., Rogers, Christopher D.F., 2020. Exploring the distances people walk to access public transport. Transp. Rev. 40 (2), 160–182.
Vansteenwegen, Pieter, Melis, Lissa, Aktaş, Dilay, Montenegro, Bryan David Galarza, Vieira, Fábio Sartori, Sörensen, Kenneth, 2022. A survey on demand-responsive public bus systems. Transp. Res. C 137, 103573.
Varela, Juan Manuel Lorenzo, Börjesson, Maria, Daly, Andrew, 2018. Public transport: One mode or several? Transp. Res. A 113, 137–156.
Wallar, Alex, Van Der Zee, Menno, Alonso-Mora, Javier, Rus, Daniela, 2018. Vehicle rebalancing for mobility-on-demand systems with ride-sharing. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 4539–4546.
Wang, Zifan, Hyland, Michael F, Bahk, Younghun, Sarma, Navjyoth JS, 2022. On optimizing shared-ride mobility services with walking legs. arXiv preprint arXiv:2201.12639.
Will, Marie-Eve, Munshi, Talat, Cornet, Yannick, 2020. Measuring road space consumption by transport modes. J. Transp. Land Use 13 (1), 651–669.
Wu, Xiatian, MacKenzie, Don, 2021. Assessing the VMT effect of ridesourcing services in the US. Transp. Res. D 94, 102816.
Yan, Xiang, Levine, Jonathan, Zhao, Xilei, 2019. Integrating ridesourcing services with public transit: An evaluation of traveler responses combining revealed and stated preference data. Transp. Res. C 105, 683–696.
Yao, Rui, Bekhor, Shlomo, 2024. A ridesharing simulation model that considers dynamic supply-demand interactions. J. Intell. Transp. Syst. 28 (1), 31–53.

Ziemke, Dominik, Kaddoura, Ihab, Nagel, Kai, 2019. The MATSim Open Berlin Scenario: A multimodal agent-based transport simulation scenario based on synthetic demand modeling and open data. Procedia Comput. Sci. 151, 870–877.

Zuniga-Garcia, Natalia, Gurumurthy, Krishna Murthy, Yahia, Cesar N, Kockelman, Kara M, Machemehl, Randy B, 2022. Integrating shared mobility services with public transit in areas of low demand. J. Public Transp. 24, 100032.

Zwick, Felix, Kuehnel, Nico, Moeckel, Rolf, Axhausen, Kay W., 2021. Agent-based simulation of city-wide autonomous ride-pooling and the impact on traffic noise. Transp. Res. D 90, 102673.