Simulation of price, customer behaviour and system impact for a cost-covering automated taxi system in Zurich
Simulation of price, customer behaviour and system impact for a cost-covering automated taxi system in Zurich

Sebastian Hörl*, Felix Becker, Kay W. Axhausen

Institute for Transport Planning and Systems, ETH Zurich, Switzerland

ARTICLE INFO

Keywords:
Automated Vehicles AMoD MATSim Zurich Transport Simulation Dynamic Demand Discrete choice

ABSTRACT

Automated Mobility on Demand (AMoD) is a concept that has recently generated much discussion. In cases where large-scale adoption of an automated taxi service is anticipated, the service’s impacts may become relevant to key transport system metrics, and thus to transport planners and policy-makers as well. In light of this increasingly important question, this paper presents an agent-based transport simulation with (single passenger) AMoD. In contrast to earlier studies, all scenario data (including demand patterns, cost assumptions and customer behaviour) is obtained for one specific area, the city of Zurich, Switzerland. The simulation study fuses information from a detailed bottom-up cost analysis of mobility services in Switzerland, a specifically tailored Stated-Preferences survey about automated mobility services conducted in the canton of Zurich, and a detailed agent-based transport simulation for the city, based on MATSim. Methodologically, a comprehensive approach is presented that iteratively runs these components to derive states in which service cost, waiting times and demand are in equilibrium for a cost-covering AMoD operator with predefined fleet size. For Zurich, several cases are examined, with 4,000 AMoD vehicles leading to the maximum demand of around 150,000 requests per day that can be attracted by the system. Within these parameters, the simulation results show that customers are willing to accept average waiting times of around 4 min at a price of 0.75 CHF/km. Further cost-covering cases with lower demand are presented, where either smaller fleet sizes lead to higher waiting times, or larger fleet sizes lead to higher costs. While our simulations indicate that an AMoD system in Zurich can bring benefits to the users, they show that the system impact is largely negative. Caused by modal shifts, our simulations show an increase of driven distance of up to 100%. All examined fleet configurations of the unregulated, cost-covering, single-passenger, door-to-door AMoD service are found to be highly counter-productive on a path towards a more shared and active transport system. Accordingly, policy recommendations for regulation are discussed.

1. Introduction

Automated vehicles have been widely discussed in society and research recently. While technology is developing quickly, questions related to planning and policy-making become ever more important (Milakis et al., 2017). Their expected positive effects include...
increased travel comfort as people can perform activities other than driving (Pudane et al., 2018; Wadud and Huda, 2019). More mobility is expected to emerge for various user groups such as the elderly and children because a missing driving license or reduced alertness will no longer be obstacles. Accident rates are also predicted to drop as vehicles have shorter reaction times than human drivers and are not prone to fatigue (Winkle, 2016). A large share of automated vehicles may free up space in the urban environment, because vehicles can park far away or in hidden, compact facilities (Nourinejad et al., 2018). Efficient energy use may make vehicle fleets ecologically friendly (Wadud et al., 2016) and efficient sharing strategies, within or among households, could reduce the total number of vehicles required.

Therefore, automated mobility could be well suited for ride-hailing services (Basu et al., 2018; Oke et al., 2020; Nahmias-Biran et al., 2020). In an Automated Mobility on Demand (AMoD) service, customers would call an automated taxi and be picked up at a predefined point. They would then be driven to their destination, potentially sharing the ride with others on the way, and dropped off. Such a system has the potential to decrease the number of vehicles in the city, further enabling city planners to re-allocate urban space (Cugurullo et al., 2020).

However, automated mobility also comes with problems. Besides ethical considerations about their behavior in conflict situations, legal liability issues and challenges in cyber-security, problems with rebound effects and induced demand remain (Schoitsch et al., 2016; Kim, 2017; Lohmann, 2016). If travelling becomes so comfortable and easy that more trips take place than before, would automated vehicles create even more congestion on the roads (Meyer et al., 2017)? Would they attract demand from aggregated public transport and degrade its service level? Additionally, automated vehicles, especially when operated as a fleet service, come with empty rides, as is the case for today’s taxis, but potentially on a much larger scale. While studies predict up to two-thirds less cars on the roads (Spieser et al., 2014) with a large Automated Mobility on Demand system, what are the societal and environmental gains (Taiebat et al., 2018; Williams et al., 2020; Wadud et al., 2016) if such a fleet covers more distance than all private cars today?.

Such questions are inherently connected to the attractiveness of an AMoD service. As has been shown, price and waiting times (service level) are crucial drivers of the decision to opt for, or against, an (automated) on-demand service, as is discussed in the literature review of Becker and Axhausen (2017) or in (Krueger et al., 2016; Lavieri and Bhat, 2019; Steck et al., 2018). Likewise, waiting times are strongly connected to the demand level, i.e. how many customers use the system, driving up waiting times. Furthermore, service cost strongly depends on how much distance the fleet vehicles drive with a paying customer and how empty they are. This ratio, in turn, affects operators’ incomes and, consequently, prices requested from customers. Hence, the problem of understanding utilization and impact of an AMoD service (at a given fleet size) emerges from a complex interplay of customer preferences and fleet management.

In this paper, we report on a comprehensive treatment of these interactions, presenting a simulation of individual decision-making customers, who interact with a fleet of automated vehicles. In our approach, we show how demand drives prices and waiting times of a (single passenger) AMoD system and how, in turn, customers dynamically react to those variables. One major assumption in our study is that prices are set such that operating costs of the service are covered entirely. Finally, we report on system states in which prices, waiting times and travel demand are in equilibrium. While conceptually similar simulation studies exist (see below), we present a consistent set-up based on user preferences and vehicle costs specifically obtained for one city, thus rendering a consistent picture of a potential future automated mobility system for Zurich or other cities.

The paper is structured as follows. Section 2 gives an overview of existing studies and how methodology has evolved. In Section 3, we present the building blocks of our proposed dynamic demand simulation. Key results of future Zurich mobility scenarios are provided in Section 4, followed by a discussion of our method and planning implications. Section 6 concludes the paper.

2. Background

While extensive comparative literature reviews on system-wide simulation studies of automated mobility are available (see, e.g., Gurumurthy et al. (2019a); Fernestål and Kristoffersson (2019); Jing et al. (2020); Narayanan et al. (2020)), the following overview will focus on the demand and supply interplay in available studies.

Fagnant and Kockelman (2014) present one of the first simulation studies with system impact in mind. They impose a static, artificially generated demand on a grid network to study an AMoD system’s impact on waiting times and empty miles. A similar approach is followed by Zhang et al. (2015). Chen et al. (2016); Fagnant and Kockelman (2018) extend the framework with infrastructure placement for charging facilities and dynamic ride sharing. The demand is made more realistic by generating individual trips,
based on origin–destination (OD) flows for Austin. While previous studies use a custom-made simulator, Liu et al. (2017), Loeb et al. (2018) present a MATSim-based simulation for Austin in a trip-based, static setting, where fleet outcomes do not feed back into demand. The methodology is based on Boesch et al. (2016), who applies it in one of the first studies in the Swiss context.

Just as the studies mentioned before, others exist based on a static (one-shot) demand, such as (Martinez and Viegas, 2017) for Lisbon, (Javanshour et al., 2019) for Melbourne, (Lokhandwala et al., 2018) based on New York taxi data, (Poullhe and Berrada, 2020) for a campus application in France and (Balac et al., 2020) for an automated on-demand public transport system in Zurich. What those static-demand simulations have in common is that fleet outcomes, such as waiting times or price, do not influence demand. Rather, it is imposed statically; the studies report on fleet sizes required to achieve certain acceptable waiting times (see Fig. 1a).

Additional behavioral realism is added by simulations allowing travellers to make dynamic decisions. As one of the first papers with active decision makers, Chen and Kockelman (2016) use a multinomial logit model with mode-independent VOTs and price sensitivities to test different pricing schemes. Furthermore, customers react to waiting times by first assuming a 2.5 min waiting time for the trip, which increases in every time step, eventually making the customer switch to either his or her car, or public transport. A similar logic is applied in a campus simulation for Delft (Scheltjes et al., 2017) and in a simulation for Ann Arbor, Michigan (Liu et al., 2018), where waiting times are communicated per trip at the time of request. Further studies extend the modeling tool capabilities by simulating detailed customer-operator interactions of request, quote, dispatching and potential drop-out (Dandl et al., 2019) and uncertainty about the arrival time of customers (Hyland and Mahmassani, 2020).

While the simulations above assume bookings on short notice (similar to the use of ride-hailing services today), a range of studies looks at the wider scope when AMoD would be available as a transport mode for regular trips. Here, the focus is shifted to customers’ longer-term planning which requires making them aware of the offered service level. Levin et al. (2017) points out that simulations with iterative structure help increase the level of realism; waiting times, travel times and congestion can go into equilibrium with demand. Such closed-loop simulation studies usually apply a demand model, followed by a simulation of the transport system and a subsequent feedback of information to the demand model. Studies that follow this pattern are presented by Wen et al. (2018), Liu et al. (2019a), all of which make use of predefined pricing schemes.

Those studies were still based on individual trips, rather than interconnected and constrained daily activity patterns. Simulations that take full-day activity chains into account are performed with the SimMobility framework for Singapore (Azevedo et al., 2016; Le et al., 2019). Their model goes one step further: not only mode decisions are dynamic, but also generation of daily activity patterns is based on increases or decreases of accessibility in the transport system (Oh et al., 2020a; Oh et al., 2020b; Nahmias-Biran et al., 2020). Gurumurthy et al. (2020a) run simulations of an automated taxi system for the Chicago region with trips being generated by a discrete choice model based on consistent activity chains, but with fixed service fares.

While the papers presented above directly relate to the use of automated vehicles, the concept of shared fleets for passenger transport has been covered by a large body of literature. While agent-based simulation studies exist (e.g. Martinez et al., 2015) most publications are rooted in optimization and operations research. They, therefore, represent a fruitful source of optimization algorithms that are used in the simulations above and referred to as “dial-a-ride” algorithms (DARP). A good overview of such algorithms give Molenbruch et al. (2017), who note that in the standard problem, operational costs are minimized, subject to full demand satisfaction and service level requirements”, which relates closely to our study presented below. Further classifications and overviews of algorithms are given by Mourad et al. (2019) and Wang and Yang (2019) (from an economic perspective). As the relevant problems scale strongly with the number of vehicles and requests, especially when ride-pooling is considered, Tafreshian et al. (2020) point out the frequent use of rolling horizon approaches in which complex optimization problems are solved for a limited number of future time steps to preserve applicability to their respective use cases. Generally, optimization-based approaches have been applied to limited real-world scenarios such as Amirgholy and Gonzales (2016) who only look at a peak hour service, or Liang et al. (2016) who look at a transit feeder service with only 40 vehicles. Others make use of downsizing, e.g. Ma et al. (2017) who consider a limited demand of 1%-2% of New York’s taxi trips, or Liang et al. (2018) who extend their former use case to the whole city of Delft with a down-scaled demand served by at most 25 vehicles (representing 500 in reality). Later the study is extended with congestion and dynamic travel times (Liang et al., 2020). Policy-wise transit integration remains an important topic in this line of research, e.g. by Levin et al. (2019) and Ma et al. (2019) who look at optimal configurations to reduce total travel time by passengers. While the mentioned studies mainly consider demand by serving or rejecting trips based on penalties from the customer perspective, Liu et al. (2019a) present a framework of running an optimization problem for fleet optimization iteratively with a discrete choice model, which comes closest to the modeling set-up that we present in this paper.

It should be mentioned that ride-hailing has recently seen an increase in more macroscopic and economic modeling, e.g. by Daganzo and Ouyang (2019), Daganzo et al. (2020) where analytical models are applied to estimate required fleet sizes and service criteria for taxi and ride-pooling systems. Similarly, Ke et al. (2020) use analytical models to find operator and social optima for single-passenger and multi-passenger services, indicating that ride-pooling may be both the more equal and profitable system.

Given these different approaches it is interesting to put into perspective why agent-based models are frequently used to study the impact of automated vehicles. One explanation may be the emergence and maturation of powerful agent-based simulation frameworks such as MATSim (see below), SimMobility, or POLARIS at the same time when interest in automated mobility was rising, starting from the early 2010s. In comparison to macroscopic or more aggregated approaches, agent-based models clearly offer a higher level of detail, because movements of individual entities such as travellers and vehicles can be tracked individually and each of them can exhibit distinct attributes. However, run times often do not permit series of many simulation configurations and extensive sensitivity analyses. In comparison to common optimization approaches the aspects of complexity and modularity may be mentioned.

In terms of complexity agent-based models are designed to test strategies and changes to the transport system “as if” they would happen in reality. While optimization-based studies often need to model explicitly numerous influences such as congestion and waiting
times, partly with strong assumptions, these effects emerge from simulation dynamics in agent-based models, where on a time step by time-step based fashion capacities of vehicles, roads, and other entities can be checked and the interaction of agents are modeled with high granularity. This leads to complex behaviour, which, on the contrary, may not be fully explainable exactly because it is emerging from many small interactions. Complex agent-based models are therefore prone to showing multiple effects at once, while it is easier to isolate effects in more rigorous mathematical models.

Related to the complexity is the modularity of the common agent-based transport modeling software. While some studies may only be interested in car traffic, mode choice, analysis of emission analysis, congestion pricing, fleet management, the frameworks offer these functionalities in a modular way, where certain analysis and concepts can be activated and deactivated. Studying the impact of a fleet control algorithms is therefore often only part of a much larger system, or can become part of one in future research. Usually, it is easy to connect modules, also because the structure of agent-based models is intuitively understandable, compared to mathematical optimization models which require a certain amount of mathematical understanding and theoretical background. They, therefore, represent a valuable transport planning methodology that is close to reality, at the cost of a less rigorous and formal definition than other approaches.

The MATSim (Horni et al., 2016) framework is used by the majority of simulation studies related to automated mobility systems (Jing et al., 2020). While a detailed description of the model will be given below, previous efforts of simulating AMoD services are presented here. Initially, Bischoff and Maciejewski (2016) propose a static demand simulation, based on taxi data for Berlin (and later Barcelona, Maciejewski et al. (2016)) without dynamic decision-making. Later, the authors extend the simulation with detailed demand patterns, using individual daily activity patterns and socio-demographic attributes of Berlin residents (Ziemke et al., 2019) to study the effect of automated vehicles’ capacity improvements (Maciejewski and Bischoff, 2018). Yet, demand is still static, based on the trips of their synthetic traveller population. Horl (2017) applies the framework first in a dynamic context for an artificial test scenario of Sioux Falls, which is later extended for ride-pooling services (Wang et al., 2018). In these simulations, service characteristics, such as waiting times, directly influence the decision-making of the agents. Simoni et al. (2019) and Gurumurthy et al. (2019b) present simulations for Austin that follow the same idea. Kamel et al. (2019) apply MATSim in a first study on potential demand in Paris, and later, in the French city of Rouen (Vosooghi et al., 2019a; Vosooghi et al., 2019b). In those studies, MATSim scoring functions are extended by specific service-related attributes that have been estimated from a survey using discrete choice models. Related simulations have been performed for Budapest (Hamadneh et al., 2019; Ortega et al., 2020) and Greenwich (Segui-Gasco et al., 2019) in combination with an external fleet control component.

These studies were based on the MATSim-specific co-evolutionary algorithm for decision-making, making them suitable only for scenario-based analysis where preferences for a future AMoD transport mode must be defined relative to the existing modes of transport. To date, no consistent way of translating the result of survey based choice models (including underlying sensitivities) to MATSim scoring parameters has been proposed. Horl et al. (2019a), therefore, explore a method to directly integrate survey-based discrete choice models into MATSim (while dropping evident advantages of MATSim’s standard scoring-based approach, see discussion). A first application of this extension was presented by Horl et al. (2019b), where a fleet of automated taxis is simulated in Paris. The paper is also the first in MATSim literature where customer decisions depend on both waiting time and price, which both change iteratively based on fleet utilization.

Fig. 1 shows the simulation dynamics evolution of the literature: first, (a) studies where defining a static demand (e.g. all car trips in a region) and testing at which AMoD fleet size a certain waiting time criterion would be reached. As shown in (b), this translates into a static demand assumption (orange), where fleet sizes and resulting travel times have no influence on the demand. However, a low level of service should lead to low demand, so it was tackled by the next generation of simulations (blue). Most studies so far assume static prices based on taxi data, or propose fixed price structures in their scenario analysis. In the present study, based on a detailed cost analysis for Switzerland, we close the loop from the operator-side by imposing an (at least) cost-covering price on the customers. Hence, our goal is to arrive at a demand curve (green) with a clear maximum at a specific “maximum demand” fleet size.

Based on these considerations, we can summarize the present paper’s contribution as follows:

- While most studies assume fixed prices, we adapt prices dynamically to offer a cost-covering AMoD service based on a realistic cost model for Switzerland.
- We furthermore adapt customer-decisions dynamically, based on a survey in the canton of Zurich, specifically designed to capture residents’ attitudes towards AMoD, while most studies assume preferences relative to existing modes.
- We simulate full-day activity patterns based on a detailed synthetic population for Zurich, which restricts the attractiveness of the service for certain complex mobility patterns, but also captures influences of attributes like car ownership.

As far as the authors know, we are therefore presenting the first study in which (as of today) realistic use-case-specific preferences are combined with a detailed case-specific cost structure and a detailed synthetic population, leading to a consistent estimate of potential AMoD demand. While previous studies assume predefined pricing schemes, we analyze, based on our case-specific cost model, a cost-covering operator.

3. Methodology

We propose a dynamic demand simulation with three major components, visualized in Fig. 2. The transport system, including an AMoD service, is simulated in the mobility simulation. As described in detail below, we use a mesoscopic simulator, able to track decisions and movements of a large number of travellers, represented as agents. They interact with each other and with the vehicle agents
of an automated vehicle service, which is simulated in detail. From the full day simulations, we can measure a number of metrics, like distance driven with a customer, or the vehicle fleet’s empty distance. Those metrics are fed into the cost calculator component, based on data for Zurich, that can estimate the minimum price that an operator would need to ask from customers to sustain a full cost-covering service. This price, along with other choice attributes that can be measured from the simulation, is fed into a discrete choice model where individual travellers - with their individual mobility patterns - decide which mode of transport to use for each trip.

By running this loop of models iteratively, we can analyze the complex interplay between supply and demand in an AMoD system: Initially, a price and fleet size is fixed for the service, which motivates a certain number of travellers to use it. As more travellers use the service, waiting times become higher. Thus, attractiveness decreases, along with the likelihood of choosing this mode of transport. Eventually, the system stabilizes in an equilibrium where prices, waiting times, and demand are in a consistent state. These values are then analyzed for different fleet sizes. Secondary metrics can be measured, like the ratio of empty distance, changes in mode shares, total distance driven in the system, etc. It is thus possible to quantify an AMoD operator’s systemic impact on a city, in this case Zurich.

The three components, (a) cost calculator, (b) discrete choice model, and (c) mobility simulation will be described in detail in the following sections.

### 3.1. Cost structure model

The cost model used in this study was developed specifically for the Canton of Zurich in Switzerland (Bösch et al., 2018). It is a bottom-up study of cost components leading to final per passenger-kilometer (pkm) and vehicle-kilometer (vkm) costs of various forms of mobility, from private vehicle ownership to taxis, trains, and buses. Full costs are thus derived from investment costs, cleaning costs, maintenance costs, vehicle management costs, cost of capital, profit margin and others.

While the cost study considers urban and non-urban environments, only urban results are relevant to this study. The study finds for today’s usage patterns that urban buses could see a cost reduction from around 0.53 CHF/pkm to 0.24 CHF/pkm if they were to be automated. In contrast, costs for private vehicle ownership would stay about the same in Switzerland as today, at around 0.50 CHF/pkm. The largest decrease, though, would be for taxi services. The cost study predicts that taxi service costs in Switzerland would drop from around 2.73 CHF/pkm to only 0.41 CHF/pkm, which would make them highly competitive with private cars. It is, therefore, reasonable to assume that private operators will offer these services\(^1\).

It should be noted that costs in Bösch et al. (2018) are based on best-guess predictions of fleet utilization and empty distances, which strongly affect the service costs. In the present study, we, therefore, feed the cost model with measured values from the agent-based transport simulation. Prices then reflect actual simulated fleet usage rather than being fixed, which would distort the results. Costs in (Bösch et al., 2018) are assigned to three reference quantities: per kilometer, per trip (cleaning) and per vehicle. Total fleet cost for the operator can be described by:

\[
C_{\text{Fleet}} = c_{\text{perDistance}} \cdot d_{\text{fleetDistance}} + c_{\text{perTrip}} \cdot n_{\text{numberOfTrips}} + c_{\text{perVehicle}} \cdot n_{\text{fleetSize}}
\]  \(1\)

Here, \(d_{\text{fleetDistance}}\) describes the total distance driven by the fleet (including distance covered with and without passengers), \(n_{\text{numberOfTrips}}\) describes the number of customer trips performed by the fleet (i.e. the number of served requests), and \(n_{\text{fleetSize}}\) describes the number of vehicles in the fleet (which will be defined as a fixed simulation parameter).

From Bösch et al. (2018) we derive the following costs per unit:

- \(c_{\text{perDistance}} = 0.098\) CHF/vkm
- \(c_{\text{perTrip}} = 0.375\) CHF
- \(c_{\text{perVehicle}} = 33.30\) CHF (per day)

In our simulations, we consider a service where prices asked from customers are adapted such that the operator cost is covered

\(^1\) For costs in other cities worldwide see Becker et al. (2020).
entirely. The price structure consists of a base fare \( b \), which is a fixed simulation parameter, and a variable distance fare \( p_{\text{AMoD}} \) which is adapted given the utilization of the fleet. Given the fixed base fare \( b \) and the inputs to the cost model from Eq. 1, we calculate the price \( p \) as follows:

\[
p_{\text{AMoD}} = \max \left\{ 0, \frac{C_{\text{Fleet}} - b \cdot n_{\text{NumberOfTrips}}}{d_{\text{CustomerDistance}}} \right\} \quad \text{in} \quad \text{CHF} / \text{pkm}
\]

Here, \( d_{\text{CustomerDistance}} \) denotes the distance that is covered with a paying customer (i.e. excluding empty distance). The nominator of the fraction describes the remaining cost if revenues from the base fare are deducted from the total operator cost. Dividing this remaining cost by the customer distance gives the required distance fare to cover the operator cost. Note that this fare can become negative when the base fare revenues exceed the actual operator cost. In that case, we require that the distance fare is set to zero, effectively leading to profit for the operator. However, in most cases we consider below, the cost model operates in the non-profit zone.

### 3.2. Discrete choice-based demand estimator

To gain insights on how people would use an AMoD service in Zurich, a survey was performed in the canton with 343 respondents. This survey was conducted in two phases; in the first, respondents provided socio-demographic information and two regular trips - below and above 50 km, respectively, along with the mode of transport they would normally choose. In a second phase, mode choice experiments were presented that included conventional and automated modes. Each respondent faced 24 choice situations. Conventional motorized modes included public transport and private car; active modes included walk and bike. For automated modes, private automated cars, individual and pooled automated taxis, and automated feeders to train stations were considered.

For all alternatives, including the AMoD (individual automated taxi), different price and waiting time levels were presented to respondents to understand their preferences for such a service. For the purpose of this study, a multinomial logit model including the conventional modes as well as AMoD taxis is estimated. Private automated vehicles and automated pooled taxis are therefore excluded. For a detailed survey description and a more comprehensive model formulation, which includes further socio-demographics and attitudes of the respondent’s, readers are referred to Hörl et al. (2019c).

The data is weighted according to socio-demographic and trip attributes of the Zurich region. Reference values are derived from the national household travel survey (BFS and ARE, 2018); trips starting or ending inside the study area are selected. For re-weighting, we choose Iterative Proportional Fitting with household-level attributes number of cars and income, person-level attributes age and type of public transport subscription, trip-level attributes mode and Euclidean distance.

The model is defined by utilities for the modes car, public transport, bicycle, walking and AMoD.

The utility for car is defined by the equation

\[
\begin{align*}
\mu_{\text{car}} &= \beta_{\text{ASC.car}} + \beta_{\text{inVehicleTime.car}} \cdot \xi_{\text{TD}} \cdot x_{\text{inVehicleTime.car}} + \beta_{\text{work.car}} \cdot x_{\text{work}} + \beta_{\text{city.car}} \cdot x_{\text{city}} + \beta_{\text{cost}} \cdot \xi_{\text{CD}} \cdot \xi_{\text{CI}} \cdot x_{\text{cost.car}}
\end{align*}
\]

Fig. 3. Zurich, Switzerland with the AMoD operating area.
with $\beta$ describing the main model parameters to estimate and $x$ the trip-level attributes. The $\xi$ describe elasticities as defined below.

The attribute $x_{work}$ defines whether the trip originates or ends at a work activity; attribute $x_{city}$ describes whether the trip starts or ends inside the city area of Zurich (Fig. 3).

For public transport, the following equation has been defined:

$$u_{pt} = \beta_{ASC, pt} + \beta_{inVehicleTime, train} \cdot \xi_{TD} \cdot x_{inVehicleTime, train} + \beta_{inVehicleTime, other} \cdot \xi_{TD} \cdot x_{inVehicleTime, other} + \beta_{inVehicleTime, feeder} \cdot x_{inVehicleTime, feeder} + \beta_{transferTime, pt} \cdot x_{transferTime, pt} + \beta_{accessEgressTime, pt} \cdot x_{accessEgressTime, pt} + \beta_{headway, pt} \cdot x_{headway, pt} + \sum_{G} \beta_{ptQuality, G} \cdot x_{ptQuality, G}$$

Travel time for public transport is defined such that different public transport modes are taken into account. Attribute $x_{inVehicleTime, train}$ determines how much time the traveler spends in a train. If the connection additionally contains stages of other modes, such as buses or trams, travel time in these vehicles is considered as $x_{inVehicleTime, feeder}$, while $x_{inVehicleTime, other}$ is zero. Only if there is no train stage on the chosen route, travel time in buses, trams or ferries is considered as $x_{inVehicleTime, other}$ while $x_{inVehicleTime, feeder}$ is set to zero.

The attribute $x_{ptQuality}$ relates to a measure defined by the Federal Office of Land Use in Switzerland that quantifies accessibility to public transport in any place in Switzerland, based on proximity to public transport stops and stations and frequency of the respective lines (ARE, 2011). It is defined on five levels from A to D and “None” with A the highest. We refer to them in the utility function through

---

**Table 1**

<table>
<thead>
<tr>
<th>Mode choice model.</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Rob. std. err.</th>
<th>Rob. t-ratio(0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private car</td>
<td>$\beta_{ASC, car}$</td>
<td>0.224</td>
<td>0.698</td>
<td>0.320</td>
</tr>
<tr>
<td></td>
<td>$\beta_{inVehicleTime, car}$</td>
<td>-0.019</td>
<td>0.010</td>
<td>-1.880</td>
</tr>
<tr>
<td></td>
<td>$\beta_{work, car}$</td>
<td>-1.161</td>
<td>0.515</td>
<td>-2.250</td>
</tr>
<tr>
<td></td>
<td>$\beta_{city, car}$</td>
<td>-0.459</td>
<td>0.621</td>
<td>-0.740</td>
</tr>
<tr>
<td>Public transport</td>
<td>$\beta_{ASC, pt}$</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$\beta_{inVehicleTime, feeder}$</td>
<td>-0.045</td>
<td>0.029</td>
<td>-1.550</td>
</tr>
<tr>
<td></td>
<td>$\beta_{inVehicleTime, other}$</td>
<td>-0.012</td>
<td>0.024</td>
<td>-0.510</td>
</tr>
<tr>
<td></td>
<td>$\beta_{inVehicleTime, train}$</td>
<td>-0.007</td>
<td>0.006</td>
<td>-1.220</td>
</tr>
<tr>
<td></td>
<td>$\beta_{transferTime, pt}$</td>
<td>-0.012</td>
<td>0.020</td>
<td>-0.610</td>
</tr>
<tr>
<td></td>
<td>$\beta_{accessEgressTime, pt}$</td>
<td>-0.014</td>
<td>0.023</td>
<td>-0.620</td>
</tr>
<tr>
<td></td>
<td>$\beta_{headway, pt}$</td>
<td>-0.030</td>
<td>0.018</td>
<td>-1.670</td>
</tr>
<tr>
<td></td>
<td>$\beta_{ptQuality, B}$</td>
<td>-1.744</td>
<td>0.578</td>
<td>-3.020</td>
</tr>
<tr>
<td></td>
<td>$\beta_{ptQuality, C}$</td>
<td>-1.641</td>
<td>0.609</td>
<td>-2.700</td>
</tr>
<tr>
<td></td>
<td>$\beta_{ptQuality, D}$</td>
<td>-0.965</td>
<td>0.680</td>
<td>-1.420</td>
</tr>
<tr>
<td></td>
<td>$\beta_{ptQuality, None}$</td>
<td>-1.089</td>
<td>0.911</td>
<td>-1.190</td>
</tr>
<tr>
<td>Bicycle</td>
<td>$\beta_{ASC, bike}$</td>
<td>0.152</td>
<td>0.769</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td>$\beta_{crossTime, bike}$</td>
<td>-0.126</td>
<td>0.071</td>
<td>-1.780</td>
</tr>
<tr>
<td></td>
<td>$\beta_{headAge, bike}$</td>
<td>-2.659</td>
<td>0.974</td>
<td>-2.730</td>
</tr>
<tr>
<td>Walking</td>
<td>$\beta_{ASC, walk}$</td>
<td>0.590</td>
<td>0.879</td>
<td>0.670</td>
</tr>
<tr>
<td></td>
<td>$\beta_{crossTime, walk}$</td>
<td>-0.046</td>
<td>0.086</td>
<td>-0.530</td>
</tr>
<tr>
<td>AMoD</td>
<td>$\beta_{ASC, AMoD}$</td>
<td>-0.061</td>
<td>0.747</td>
<td>-0.080</td>
</tr>
<tr>
<td></td>
<td>$\beta_{inVehicleTime, AMoD}$</td>
<td>-0.015</td>
<td>0.010</td>
<td>-1.530</td>
</tr>
<tr>
<td></td>
<td>$\beta_{waitingTime, AMoD}$</td>
<td>-0.093</td>
<td>0.100</td>
<td>-0.930</td>
</tr>
<tr>
<td></td>
<td>$\beta_{work, AMoD}$</td>
<td>-1.938</td>
<td>0.610</td>
<td>-3.180</td>
</tr>
<tr>
<td>Others</td>
<td>$\beta_{cost}$</td>
<td>-0.089</td>
<td>0.021</td>
<td>-4.270</td>
</tr>
<tr>
<td></td>
<td>$\beta_{SP}$</td>
<td>0.942</td>
<td>0.155</td>
<td>6.060</td>
</tr>
<tr>
<td></td>
<td>$\xi_{C1}$</td>
<td>-0.817</td>
<td>0.352</td>
<td>-2.320</td>
</tr>
<tr>
<td></td>
<td>$\xi_{CD}$</td>
<td>-0.221</td>
<td>0.078</td>
<td>-2.840</td>
</tr>
<tr>
<td></td>
<td>$\xi_{TD}$</td>
<td>0.115</td>
<td>0.403</td>
<td>0.290</td>
</tr>
<tr>
<td></td>
<td>$\theta_{referenceDistance}$</td>
<td>39.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$\theta_{referenceIncome}$</td>
<td>12.260</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

| LL(0)          | -3683.030 |
| LL(final)      | -2252.310 |
| Adj.Rho-square | 0.381     |
| #Individuals   | 343       |
| #Parameters    | 28        |
| #Observations  | 5159      |
$G \in \{ B, C, D, \text{None} \}$.

Utilities for cycling and walking are defined as

$$u_{\text{bicycle}} = \beta_{ASC, \text{bicycle}} + \beta_{\text{travelTime}, \text{bicycle}} \cdot \xi_{TD} \cdot x_{\text{travelTime}, \text{bicycle}} + \beta_{\text{highAge}, \text{bicycle}} \cdot \left[ \text{age} \geq 60 \right]$$

(5)

and

$$u_{\text{walk}} = \beta_{ASC, \text{walk}} + \beta_{\text{travelTime}, \text{walk}} \cdot \xi_{TD} \cdot x_{\text{travelTime}, \text{walk}}$$

(6)

Variables $a$ refer to agent-level attributes, and, for the case of the bicycle mode, specifically to the age of each agent. Agent’s age is not included in the walk alternative, since its parameter does not have a significant influence.

Elasticities of Euclidean distance on travel time $\xi_{TD}$ and on cost $\xi_{CD}$ are defined as

$$\xi_{TD} = \left( \frac{x_{\text{EuclideanDistance}}}{\theta_{\text{ReferenceDistance}}} \right)^{\lambda_{TD}}$$

and

$$\xi_{CD} = \left( \frac{x_{\text{EuclideanDistance}}}{\theta_{\text{ReferenceDistance}}} \right)^{\lambda_{CD}}$$

(7)

with $\lambda$ describing additional model parameters that must be estimated. Reference distance is set as 39 km, the observed sample average.

Elasticity of household income on cost $\xi_{CI}$ is defined as

$$\xi_{CI} = \left( \frac{\text{householdIncome}}{\theta_{\text{ReferenceIncome}}} \right)^{\lambda_{CI}}$$

(8)

Reference household income amounts to 12,260 CHF, the observed sample average.

Costs for the car mode are defined as 0.27 CHF/km, based on driven distance along the planned route. Costs for public transport are based on additional agent-level attributes, such as subscription ownership for the GA, which is a commonly used, annually billed subscription giving access to the entire Swiss public transport system.

While the model presented so far refers to modes available in the baseline case, an additional AMoD utility can be defined based on the survey:

$$u_{\text{AMoD}} = \beta_{ASC, \text{AMoD}} + \beta_{\text{inVehicleTime}, \text{AMoD}} \cdot \xi_{TD} \cdot x_{\text{inVehicleTime}, \text{AMoD}} + \beta_{\text{accessEgressTime}, \text{AMoD}} \cdot x_{\text{accessEgressTime}, \text{AMoD}} + \beta_{\text{waitingTime}, \text{AMoD}} \cdot x_{\text{waitingTime}, \text{AMoD}} + \beta_{\text{work}, \text{AMoD}} \cdot x_{\text{work}} + \beta_{\text{highAge}, \text{AMoD}} \cdot \left[ \text{age} \geq 60 \right] + \beta_{\text{cost}} \cdot \xi_{CD} \cdot \xi_{CI} \cdot x_{\text{cost}, \text{AMoD}}$$

(9)

Estimated model parameters are documented in Table 1. The model was estimated with the R-package Apollo (Hess and Palma, 2019b; Hess and Palma, 2019a).

For the scope of this research, it is interesting to compare the value of travel time savings (VTTS), defined as

$$VTTS = \frac{\partial u_m}{\partial x_{\text{travelTime}, m}} / \frac{\partial u_m}{\partial x_{\text{cost}, m}}$$

(10)

of various modes to each other. For the reference income, respondents are willing to pay less to reduce the travel time in an AMoD service compared to their private cars (Fig. 4). This indicates that they expect it to be more comfortable. Similarly, Steck et al. (2018) observed that respondents from Germany are also observed slightly lower VTTS for an AMoD service compared to the private car (6.46 EUR/h vs. 7.22 EUR/h).

In the model presented here, the VTTS are comparable to the bus, but considerably higher than for trains. Since the travel times of an AMoD service are very competitive, this shows that there is a substantial market potential, also from a behavioral perspective. One should, however, take notice that the travel time parameters are not significantly different from each other in this model.

However, Fig. 4 also shows the VWTS (value of transfer waiting time savings) defined as
and a analogously derived value for the valuation of the transit connection’s headway. Note that waiting time for public transit in the model refers to waiting time between two transit stages, hence the valuation of the headway translates more closely to the schedule delay experienced before entering an AMoD vehicle. In any case, both the valuation of transit waiting time and headway are substantially lower than valuation of waiting time for the AmoD service. This finding indicates - unsurprisingly - that a major objective for any AMOD operator should be to minimize customer waiting times.

3.3. Baseline simulation

For the simulation, the agent- and activity-based transport simulation framework MATSim (Horni et al., 2016) is used. MATSim is an agent-based transport simulation framework, in which travelers in the real world are represented by artificial agents. Those agents have socio-demographic attributes such as age and gender, as well as a daily schedule. These consist of activities, containing information about location of those activities, their start times, durations and types. Further, those activities are connected by trips described by a specific mode of transport and a route through the transport system.

The simulation considers a 24 h day, which is simulated second by second. At each time step, agents are either in an activity (being at one specific location) or in transport (for instance, on the road network or in a public transport vehicle). Movements of road vehicles are simulated along a directed graph network representing roads of the infrastructure. Such infrastructure has capacities (for instance, limited space on roads) such that interactions between agents lead to congestion, which is simulated making use of a queue-based traffic simulation (Horni et al., 2016).

The basis of an agent-based MATSim simulation is a synthetic population of the case study area, which has the aim to represent the real population and their daily trips sufficiently well. It is generated from a range of different Swiss data sets, which are available for research purposes. Their use to generate the population is presented in Appendix A. The appendix also covers how the model is calibrated to reproduce appropriately the daily mode shares over various distance classes and the travel time distributions observed in reality. During calibration, network capacities are adjusted, the parameter \( \beta_{ASC,car} \) is adjusted to fit mode shares well, and a penalty is added to the choice model to suppress frequent use of “walking” as it was not frequently chosen in the survey, which lead to misleading predictions.

Further assumptions for the baseline model are described in detail in Appendix A:

- A perimeter of 30 km around the center of Zurich is simulated.
- With a network of around 150,000 links and over 2 million agents, we perform simulations on a 10% sample of the population to save computation time. Road capacities are scaled accordingly.
- Public transport is simulated according to the planned schedule, there is no interaction with traffic in the road network, and no crowding is considered.

The mobility simulation of MATSim is run in a loop with the discrete choice model to obtain the results. In each iteration, 5% of the agents perform new mode choice decisions for all tours and trips in their daily schedules.

3.4. AMoD simulation

A couple of extensions exist in the MATSim ecosystem to simulate dynamic transport services such as AMoD. Their common denominator is the DVRP extension of MATSim (Maciejewski et al., 2017). Unlike standard traveller agents, which follow predefined daily plans in every iterative run of the mobility simulation, DVRP allows us to control agents dynamically, second by second. A streamlined implementation of this process is provided by Horl (2017). Using this component, a dynamic fleet operator can dispatch vehicles to customers, which are picked up on arrival, transported to their destination and then dropped off. The virtual operator receives requests from traveller agents (without any pre-planning) and reacts to them by deciding which idle vehicle to send to which customer.

Using this framework, simulations have already been performed in the Zurich context in a set-up very similar to the one described here. In (Horl et al., 2019d), the authors compare different fleet operating policies in terms of how well they are able to serve the total (static) motorized mobility demand of Zurich. For the current study, the Load-Balancing Heuristic Policy first introduced by (Bischoff and Maciejewski, 2016) is chosen because Horl et al. (2019d) show that it provides a good trade-off between produced empty distance and waiting time, but, most importantly, can be executed very quickly. In brief, the control policy first determines whether it should operate in “over-supply” mode (if there are more idle vehicles than open requests) or in “under-supply” mode (the other way round) at the beginning of each decision period of ten seconds of simulated time. In over-supply mode, the algorithm loops through all requests and assigns the closest vehicle, which allows for an even service across the region with potentially long empty rides. However, in under-supply mode, the algorithm loops through all vehicles that have just become idle and assigns the closest waiting request. The idea is to decrease quickly the number of pending requests at times of high demand. In our model configuration, all issued requests must be served, although long waiting times may occur.

To integrate the AMoD simulation into our dynamic demand experiment, two additional components had to be added: functionality for predicting waiting times and functionality to calculate fleet cost and resulting prices, both to be fed into the choice model for the
Waiting times are estimated based on hexagonal zones of 500 m (outer) radius as shown in Fig. 3 and in 15 min intervals. Initially, the waiting times are set to 10 min for each time bin and zone. Afterwards, during every mobility simulation, we track when agents in each time bin and zone depart and when they enter a fleet vehicle. The difference in time of these two events is the waiting time of the trip. By calculating the mean over all waiting times that are observed for one combination of zone and time bin, we establish the estimated travel time to be fed to the discrete choice model in the next iteration of the modeling loop. In case no waiting time is observed for a combination of zone and time bin in a particular iteration, the existing value is maintained.

The difficulty of defining the price to be sent to the choice model is that initially, at zero or very low demand, prices are unrealistically high, because a cost-covering price for a fleet without customers is calculated. Therefore, we run 25 iterations with a fixed price of 0.7 CHF and only after that we use the price produced by the cost model. Preliminary experiments have shown that at this point the model produces reasonable prices given the demand and that the modeling loop is able to adapt reliably to the “shock” that is induced to the system by changing the price abruptly.

The simulations are stopped once the number of requests, the calculated distance fare, the waiting times and travel times in the network stabilize around a stochastic equilibrium. The respective procedure and criteria for convergence are described in detail in Appendix B.

To enforce realistic behavior going beyond what the choice model is able to replicate, we introduce three constraints to the choice process for the AMoD service:

- **Service area**: Only trips that start and end (origin and destination coordinates) inside the operating area (Fig. 3) are eligible for the AMoD service.
- **Minimum distance**: Trips shorter than 0.25 km (in Euclidean distance between origin and destination coordinates) cannot be performed with the AMoD mode.
- **Maximum waiting time**: Trips with a predicted waiting time of more than 15 min will not be offered AMoD as an alternative mode.

The first two constraints naturally define the service offer of the operator. In reality, it would be unlikely for such a service to dispatch vehicles for trips that only take a few seconds. The last constraint is more interesting, as it relates to the customer perspective. We assume that a customer who, at his or her usual time and place of departure, has an expected (i.e. common) waiting time of more than 15 min would refrain from using the service at all. Similarly, an operator may not even provide a trip beyond such a limit. Effectively, this constraint can therefore also be understood under the notion of trip rejection, commonly used in other models.

4. Simulation

In the following we present simulation results with varying a priori defined fleet sizes (between 1,000 and 8,000 vehicles) and base fares (zero cost, 1 CHF, 2 CHF). We perform each simulation with five different random seeds. The shaded areas in the plots below display the obtained value range, while lines show the mean of those five samples. All analyses are based on trips occurring entirely inside the operating area of the AMoD service, i.e. origin and destination are within this region. All simulations are carried out on a 10% sample of the population with an accordingly scaled fleet size, while the analyses show up-scaled simulation results.
The following analysis takes on three perspectives: the operator perspective looking at operational insights from the simulations, the user perspective looking at how the service is used and perceived by the population, and, finally, the system perspective looking at changes in the overall distribution of traffic.

4.1. Operator perspective

From the operator viewpoint, we are first interested in the attractiveness of the service, quantified by the number of requests that are sent during one day. Those are strongly dependent (as of how the model is defined) on the waiting times that the system offers, and the price of the service. As the base fare is fixed in the presented experiments, the distance fare becomes the determining factor of the cost-covering service. For further discussion, it is furthermore useful to know the underlying fleet cost that defines this price.

A major cost component of the service is the fleet size. The choice of fleet size can be regarded as a mid- to long-term decisions and can not be changed easily from one day to another. Research on on-demand services therefore focuses on how those vehicles are operated, and the major metrics of interest are the total fleet distance and the share of empty distance for which vehicles drive without a customer. Lastly, it is interesting to analyze how much distance individual vehicles travel.

4.1.1. Service analysis

Fig. 5 shows simulation results for important operator metrics. Fig. 5a shows how the demand (number of requests) for the service depends on the fleet size. The curves for three different base fares show how small fleet sizes lead to low demand, followed by an increase up to a demand maximum at around 4,000 vehicles, followed by declining demand for even larger fleets. These results quantify the shape of the qualitative demand curve postulated in the paper’s introduction in Fig. 1. Regarding the effect of the fare, one can see that higher base fares systematically lead to lower demand. This is not an obvious finding as distance fares in the cost-covering service adapt according to the chosen base fare.

In Fig. 5b the achieved waiting times are shown. They are measured from simulation by considering all trips performed with the AMoD service and noting down their observed waiting time. Both the mean and the 90% quantile over the whole simulated day are reported. There is a clear trend for small fleet sizes to produce higher waiting times, while larger fleet sizes lead to lower waiting times. The almost non-existent variation of waiting times across base fares is explained by the fact that Fig. 5 shows waiting times that are accepted by the customers. It is important to remember that the choice of transport mode is defined by a mode choice model. Hence, the displayed values depend strongly on the choice behaviour of the agents. We can therefore state that at a fleet size of 4,000 vehicles the travellers in Zurich would be willing to accept on average a waiting time of 4 min, which, in 90% of the cases, is less than 7.5 min. Yet, the travellers are willing to accept higher waiting times at smaller fleet sizes, up to 10 min on average at 1,000 vehicles; and they demand lower waiting times of around 2 min at a large fleet of 8,000 vehicles. The different levels of acceptance can be explained by the different prices asked from the customers for each trip at different fleet sizes. Fig. 5c shows the distance fare that is calculated to cover the cost of operating and maintaining the fleet. It resembles directly $p_{\text{AMoD}}$ as has been defined in Eq. 2. One can see a clear influence of the base fare such that the distance fare is reduced if the base fare is increased. Only for few cases with small fleet sizes and a high base fare of 2 CHF we observe cases in which the distance fare becomes zero because the revenues from the base fare exceed the cost of the fleet.

The cost (as defined in Eq. 1) is shown in Fig. 5. At the “maximum demand” case of 4,000 vehicles without base fare, daily costs of
about 220,000 CHF are measured from simulation.

4.1.2. Distance analysis

Fig. 6 gives further insights on operational aspects by providing a more detailed analysis of the distances driven by the AMoD fleet.

In Fig. 6a the total fleet distance is shown, i.e. all driven distance by the fleet vehicles, with or without customer. The latter is presented separately in Fig. 6b. The values are measured directly from simulation by summing up all the distance covered by the simulated vehicles and tracking down whether they have a customer on board or not. It is interesting to see that the maximum total distance occurs at higher fleet sizes than the peak of empty distance, which lies around 2,000 vehicles in all presented cases.

Generally, an operator can increase profits or improve the service level by reducing empty distance in its service, e.g. by applying a more efficient fleet control strategy. It is therefore interesting to analyze how much of the driven distance is covered without a customer (and is, therefore, not generating revenue). The share of empty distance is shown in Fig. 6c. For a small fleet size of 2,000 vehicles it reaches around 35%, while a large fleet sizes can produce less than 20% of empty distance. This effect can be explained by an increased general availability of vehicles in the network.

Finally, Fig. 6d shows the distance driven per vehicle and day. It is measured by tracking the driven distance of each individual fleet vehicle and then looking at the distribution of these individual distances. In Fig. 6d it can be seen that vehicles on average drive up to 350 km per day for small fleet sizes, while their activity is low for a large fleet. In those cases, the fleet is rather resting idle to quickly serve nearby requests when they pop up. Besides the mean value, Fig. 6d also shows the maximum value of the vehicle distance distribution. For the smallest shown fleet size, the vehicle with the longest distance drives about 400 km on a single day, while the most active vehicle at a fleet size of 8,000 vehicles drives around 100 km. These values are at the upper and of the range of modern battery electric vehicles and indicate that few vehicles would likely need to recharge once during the day, while many can be recharged overnight.

4.2. User perspective

Looking at the users of the AMoD system, we are interested in how the service is used and experienced. As this analysis becomes more detailed than the operator perspective, we limit the analysis to three fleet sizes of 2,000 vehicles representing a small case, 4,000 vehicles representing the “maximum demand” case without base fare, and 6,000 vehicles representing the case of a large fleet. Also, we are interested in how the choice of the base fare affects the behaviour of the travellers.

As metrics, waiting times and prices are important from the customer perspective as they are the major factors influencing the attractiveness of the service. Based on the offered service, specific usage patterns arise which shall also be studied in the following. Of interest are the trip distances covered with the AMoD service and the range segments in which the automated taxi system becomes a strong competitor for the traditional transport modes.

4.2.1. Temporal use and waiting time analysis

Fig. 7 shows a range of time-based analyses for the selected fleet sizes. The first two graphs (a,b) show average waiting times and 90% quantile waiting times. The latter can be interpreted as a measure of reliability as they state that 90% of trips have waiting times below this value. Waiting times are measured from simulation by noting down all customer trips of the service and saving their
departure and waiting time. We then group departure times in intervals of one hour and find the mean and 90% quantile of the distribution of waiting times in each interval. Figs. 7 a and b show how the overall waiting times from the operator perspective relate to different times of the day. In general, three peaks can be identified in the morning around 8am, at noon around 12 pm, and in the evening at around 6 pm. Over the course of a day, waiting times vary more for the smaller fleet size, which can be explained by lower vehicle availability and thus lower reliability.

The patterns of increases in waiting times are reflected by Fig. 7c, which shows the number of active and waiting requests by time of day. For this analysis, we consider all customer trips during one day and probe in 10 min intervals how many requests have been sent before that time, but have not been picked up. Summing up the number of those requests leads to the number of waiting requests. Active requests are those requests that may or may not have been picked up, but for which the customer did not arrive yet at the destination. Fig. 7 clearly shows peaks of active requests at the aforementioned times. The large fleet size and the "maximum demand case" show the same patterns of active requests while demand is clearly capped in the afternoon for a small fleet size.

4.2.2. Prices

Fig. 8 shows the prices paid by the travellers for different base fares and the three selected fleet sizes (represented by different graph types). The values are obtained by looking at all AMoD trips in each simulated scenario and then calculating the price by summing up the scenario’s base fare and the in-vehicle distance multiplied by the active distance fare. This allows us to present distributions of paid prices. Cases in which only the base fare is paid (for the small fleet size and base fares of 1 CHF or 2 CHF) are not visualized in Fig. 8 as they are represented by close-to-singular distributions at the respective base fare.

Fig. 8 clearly shows the influence of an increasing base fare: the peak of the paid prices is shifted towards higher values.
Interestingly, decreasing fleet sizes lead to spikier distributions with the mode of the distribution at lower prices. This is not only an effect from the lower prices offered by the service operator, but also because longer trips are avoided due to the lower reliability of the service.

4.2.3. Trip distance analysis

Fig. 9 shows an analysis of the distances covered by trips with the AMoD service. To perform these analysis, all AMoD trips are considered from simulation and their distances are noted down. Subsequently, distance distributions and their mean can be obtained. On top, Fig. 9 shows those mean distances for three fleet sizes in combination with the different base fares. A clear pattern arises for the base fares: the higher the fare, the longer becomes the average travelled distance. This is the direct effect of pricing out many short trips that become exceedingly expensive with a higher base fare. From the case without a base fare to the one with 2 CHF the mean distance increases by about 500 meters. This increase is similar for all shown fleet sizes. Looking at one specific base fare, the mean distance slightly but systematically decreases with increasing fleet size in most cases, which can be explained by an increased vehicle availability and lower waiting times.

At the bottom of Fig. 9 the distance distributions are shown. While they look similar for fleet sizes of 4,000 and 6,000 vehicles, the demand is clearly reduced for 2,000 vehicles. For all fleet sizes one can see that the reduction of demand due to increasing base fares happens mainly at lower distances in a band up to 2.5 km. Nevertheless, trips in this range are attractive target for the automated taxi service.

4.2.4. Mode-specific distance distributions and shifts

From the user perspective, it is finally interesting to analyze which former trips the AMoD service replaces at the “maximum demand” fleet size of 4,000 vehicles. Each line in Fig. 10 represents an analysis for a traditional mode including the private car (top row), public transport (second row), and active modes (walking and cycling, third row). The left-hand side shows the baseline distance distribution for the respective mode as a dotted line together with the mode-specific distribution in the AMoD scenarios. The right-hand side shows the difference between the number of trips in the baseline scenario versus the number of trips in the AMoD scenario, i.e. the plots quantify the “loss” of trips at a certain distance.

Looking at the left-hand side, one can observe that the mode-specific distance distributions look similar for different base fares,
which, however, is also a question of scale in the presented plots. More interesting is a comparison between the three modes. While the AMoD system reduces the number of cars trips largely over the whole range of analyzed distances, reductions for public transport are more limited to a distance band between two to four kilometers. For the active modes, almost no reduction can be seen in comparison to the a priori large number of trips.

The right-hand in Fig. 10 shows the reduction in trips from a slightly different angle. Comparing private car trips and public transport it becomes evident that the distance band between two and four kilometers strongly attracts public transport users. While the peak of car trip reduction is around four kilometers and at about 2,000 trips, the peak in reduction for public transport is at around 3 kilometres with almost 4,000 trips. Fig. 10f paints a clearer picture of the influence of the AMoD fleet on the active modes. While for the private car and public transport increasing base fares do not lead to noticeable differences, there is a clear effect of stronger trip reduction with increasing base fares for the active modes. At a base fare of 2 CHF, about 13,000 active trips are replaced by AMoD, and even up to 20,000 if no base fare is used.

4.3. System perspective

Finally, the impact of the AMoD fleet on the overall transport system shall be analyzed as those results can lead to informed decisions in infrastructure planning and policy-making. In a first step, the change of the overall modal split will be examined as it is a common goal in urban transport planning to increase, for instance, the shares of public transport or active modes. Second, we will examine how the introduction of the AMoD fleet affects the use (and deterioration) of road infrastructure, but also how it can help to reduce the use of space in the city. Last, we perform a detailed analysis of congestion and road use.

4.3.1. Mode share analysis

Fig. 11 shows the generated mode shares after the automated taxi system is introduced to the transport system of Zurich. It is obtained by tracking all trips in the simulation and denoting their mode and travel distance. The total distance per mode is summed up and divided by the total distance of all trips of the simulation run. While we present hence a distance-based mode share, results on trip-based mode shares have been analyzed, but not reported here as they show the same patterns.

On top, Fig. 11a shows the achieved mode share of the AMoD service. At the “maximum demand” case of 4,000 vehicles the attracted requests represent a substantial distance-based mode share of almost 20%. The shares of the traditional modes drop accordingly. The share of the private car drops by up to six percentage points, while the public transport share even drops up to seven points. The slightest reduction happens for the active modes with only up to five points.

Interestingly, the choice of base fare has a noticeable effect on the mode share, especially for larger fleet sizes. Imposing a base fare of 2 CHF decreases the mode share of public transport by more than 1% compared to the case without base fare. This result may seem counter-intuitive at first but becomes clear in the light of the previous analyses: As shorter distances are priced out with increasing base fares, we observe a shift of demand towards longer (mean) distances. The fact that the demand stays roughly the same at a given fleet size indicates that the freed capacity from pricing out those trips is then filled up by longer ones.

4.3.2. Road distance vs. use of space

Fig. 12 finally shows two contrary findings from the system perspective. On the left, the distance driven on roads in the operating
area is presented. It is calculated by summing up all the distance driven on the roads inside the operating area by private cars and automated taxis. Additionally, we measure the distance driven by busses from the digital transit schedule and add the resulting 35,000 km per day to the distance driven per day. Fig. 12a shows the resulting values for different fleet sizes and base fares in absolute terms and compared to the baseline scenario with around 380,000 km per day. Note that the reported values, especially the relative ones, need to be interpreted with care as major components are missing, such as the distance originating from heavy freight vehicles, delivery and service trips, and tourism.

In the worst case (from a perspective of infrastructure deterioration and maintenance, and traffic), the distance driven on the roads per day increases by up to 85% for the maximum demand case. At both rather small fleets of 1,000 vehicles and large fleets of 8,000 this increase only reaches about 60% of added distance. Interestingly, we observe that a higher base fare leads to higher road usage, while some distance can be saved by running the system without any initial price.

On the right, Fig. 12b shows the number of used vehicles in the operating area during one day. This value is obtained by considering all trips in the operating area and counting how many individual private cars are used for those trips. Effectively, this means that we count the number of individual persons that drive a car at any point during the day, because cars cannot be shared in the simulation. Additionally, we add the number of fleet vehicles as defined in the respective simulation scenario. We do not consider busses in this value as only transit schedule information does not allow a straightforward calculation of this number. Fig. 12b shows the obtained number of used vehicles per day in different scenarios. The largest reduction can be seen for a fleet size of 3,000 vehicles. At that point the overall vehicle fleet is reduced by more than 13%. Smaller reductions are observed for all fleet sizes.

4.3.3. Congestion and road use

Given the substantial increase in distance driven on the roads caused by modal shifts it is interesting to investigate how this heavier use of infrastructure translates to congestion issues. To find out whether congestion increases, we have applied the following procedure. As the structure of the agents’ activity chains in our simulation does not change (only mode choice is considered), it is possible to link trips done by the same person at the same position in the daily activity chain between the baseline simulation and the AMoD cases. By filtering for all corresponding trips which are done using the private car in both the baseline and AMoD simulation, we can compare their travel times. If we can see that statistically travel times increase after the introduction of the AMoD system, we can conclude that there is more congestion. To quantify this increase, we calculate the relative increase of travel time for each obtained trip and then establish the distribution of relative changes. The mean and the 90% quantile of this distribution is shown in Fig. 13 for each of the various cases.

Fig. 13 shows that there is an increase in congestion (in terms of increasing travel time). On average, travel times increase by 0.5% to 1.5% for the examined cases. However, Figure Fig. 13 shows that in 10% of the cases travel times have increased by more than 10%. Together, those values indicate that added congestion is affecting travellers quite differently and that especially at peak times increases may be higher.

An interesting question to ask is where in the system congestion happens. Fig. 14 shows such an analysis for the specific example of
a fleet size of 4,000 vehicles without base fare. On the left-hand side the traffic volumes in the baseline simulation are shown with roads in blue being used by only 1,600 vehicles per day, while the count for roads in red exceeds, often strongly, 8,000 vehicles per day. The map reflects the traffic patterns of Zurich with the major roads clearly sticking out in red color. On the right-hand side, Fig. 14 shows the simulation with automated taxis in comparison to the baseline case. Roads colored in blue see an increase of vehicles per day by up to 20%, while roads colored in red have an increase to more than twice the amount of vehicles in the baseline simulation. Comparing these two maps renders a clear picture: While the major roads have only slight increases in vehicle use, there is a substantial impact on smaller roads in the residential areas of the city.

Based on these insights, Table 2 allows us to have a more quantified look. For the same case (4,000 vehicles and zero base fare), three analyses are presented: the share of the total distance driven in the baseline simulation by road category; the increase in distance driven on roads inside each category; and the share of the total distance driven by the AMoD fleet in each category.

First, Table 2 shows that in the baseline case most of the driven distance occurs on primary roads (mainly the ones shown in red in Fig. 14a). Second, we see a clear increase in driven distance for the lower order road types, especially for residential roads with an increase of more than 30% compared to the baseline. Last, the use of roads by the AMoD fleet is quantified. About 60% of all fleet distance is driven on tertiary or residential roads. The strong increase of total driven distance in the system in combination with the relatively small increase in congestion can therefore be explained by the fact that most of the added distance is performed on lower order roads for the purpose of picking up and dropping off customers at their trip origins and destinations.

5. Discussion

The following paragraphs provide a discussion of the results presented above. First, the results will be interpreted in general terms with a focus on how the dynamic price adaptation compares to earlier studies for automated taxi services. Second, these considerations

---

Table 2

<table>
<thead>
<tr>
<th>Road type</th>
<th>Baseline share</th>
<th>Increase/Decrease</th>
<th>AMoD share</th>
</tr>
</thead>
<tbody>
<tr>
<td>motorway</td>
<td>8.69 %</td>
<td>-0.73 %</td>
<td>0.86 %</td>
</tr>
<tr>
<td>trunk</td>
<td>3.97 %</td>
<td>-0.16 %</td>
<td>0.40 %</td>
</tr>
<tr>
<td>primary</td>
<td>33.78 %</td>
<td>+5.94 %</td>
<td>21.01 %</td>
</tr>
<tr>
<td>secondary</td>
<td>14.78 %</td>
<td>+9.69 %</td>
<td>15.89 %</td>
</tr>
<tr>
<td>tertiary</td>
<td>24.17 %</td>
<td>+11.78 %</td>
<td>30.11 %</td>
</tr>
<tr>
<td>residential</td>
<td>14.61 %</td>
<td>+31.04 %</td>
<td>31.72 %</td>
</tr>
</tbody>
</table>

The categories are based on OpenStreetMap attributes, with motorway and motorway_link combined into one category. The same scheme applies for the other main categories. The residential category also includes the types living_street and unclassified, which are also included in the simulated road network.
lead to a closer analysis regarding policy implications and recommendations for Zurich. Finally, an outlook is given how remaining questions and methodological challenges can be overcome in future developments of the framework.

5.1. General discussion

The simulation results show the behaviour postulated in the introduction: Fig. 5 shows a clear demand curve where small fleet sizes lead to low demand (because of poor service levels) and large fleet sizes also suppress demand because of high prices. The high prices result from the constraint that we require the service to be cost-covering. All presented configurations of fleet size, waiting time, price, and demand are, therefore, economically feasible, i.e. an operator providing such a service would neither make losses nor profits. Thus, the simulated operator should be seen as a public entity, for instance, the local transport agency which wants to offer a self-sustaining service without additional subsidy funding.

Under this assumption, the presented results allow such an operator (or the city, the planning authorities, ...) to choose which service should be offered: Is it desirable to have an inexpensive service with high waiting times? Or should a service with high availability be put in place? In the latter case, the prices adapt accordingly and the service becomes a mode of transport for the wealthier part of the population. It is important to note that both cases, either a cheap and slow service, or an expensive and fast service, can lead to similar outcomes in terms of the overall system impact (mode share, total driven distance, ...). Which configuration to choose becomes thus also a question of equity and accessibility to the service.

It is interesting to think of what happens if configuration parameters change slightly. At constant fleet size, increasing the price would lead to profit for the operator. However, the increased price would diminish demand, which, in turn, would free up service vehicles leading to shorter waiting times, which would attract demand. This example shows the complex interplay of the involved components, which can be studied with our presented framework. In future studies, it would be interesting to study how the equilibria align if a profit margin on the price is required by the operator. An analysis like Fig. 5 could then help the city and operators to choose the service configuration that fits the interest of both parties best.

In general, the range of our results fit well to existing research. The accepted waiting times from Fig. 5 of three to ten minutes fit well into the range of values that has been obtained by previous studies (Jing et al., 2020). In these studies, waiting times of five or ten minutes have often been defined a priori as evaluation criteria. Our results, based on dynamic decision making and our survey, show that these are indeed relevant thresholds for a moderately expensive service.

Regarding prices, it is interesting to look at which operating configurations have been found in other studies. Table 3 shows a selection of studies for which it is easy to derive a notion of service price, fleet size, and demand. For better comparison, the obtained (or defined) prices have been converted from their local currencies to units of purchasing power parity (OECD, 2020). Generally, the values fall into the same range from about 0.30 PPP/km to 1.00/km. Our results (Fig. 5, and last row in Table 3) show that a value of 1.22 PPP/km is already a high value at which demand declines in our cost-covering simulations.

The upper part of the table shows simulations in which a demand-affecting price is defined, but then the fleet is optimized to reach a certain level of service. The lower part of the table includes the present paper and (Hörl et al., 2019b) in which we used a similar simulation set-up for Paris with fleet size and attracted demand governing the price.

In the upper part, for non-adaptive prices, there is a clear correlation between increasing prices and a decreasing number of requests. Based on the set-up of the studies, fleet size decreases accordingly to get close to the waiting time threshold. While no analysis is

Table 3

Comparison of distance fares, fleet sizes and demand for related simulation studies.

<table>
<thead>
<tr>
<th>Price</th>
<th>PPPa</th>
<th>Fleet size</th>
<th>Requests</th>
<th>Waiting timeb</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Chen et al., 2016)</td>
<td>0.75 USD/km</td>
<td>0.75 PPP/km</td>
<td>137,000</td>
<td>39.05%</td>
</tr>
<tr>
<td>1.00 USD/km</td>
<td>1.00 PPP/km</td>
<td>45,500</td>
<td>14.36%</td>
<td>3 min</td>
</tr>
<tr>
<td>(Liu et al., 2017)</td>
<td>0.50 USD/km</td>
<td>0.50 PPP/km</td>
<td>17%</td>
<td>43,000</td>
</tr>
<tr>
<td>Austin</td>
<td>0.75 USD/km</td>
<td>0.75 PPP/km</td>
<td>13%</td>
<td>16,700</td>
</tr>
<tr>
<td>1.00 USD/km</td>
<td>1.00 PPP/km</td>
<td>13%</td>
<td>8,900</td>
<td>3 min</td>
</tr>
<tr>
<td>(Oh et al., 2020a)</td>
<td>75%</td>
<td>43,000</td>
<td>883,000</td>
<td>4-6 min</td>
</tr>
<tr>
<td>Singapore</td>
<td>100%</td>
<td>33,000</td>
<td>704,000</td>
<td>4-6 min</td>
</tr>
<tr>
<td>125%</td>
<td>27,000</td>
<td>580,000</td>
<td>4-6 min</td>
<td></td>
</tr>
<tr>
<td>(Hörl et al., 2019b)</td>
<td>0.25 EUR/km</td>
<td>0.34 PPP/km</td>
<td>10,000</td>
<td>800,000</td>
</tr>
<tr>
<td>Paris</td>
<td>0.31 EUR/km</td>
<td>0.42 PPP/km</td>
<td>25,000</td>
<td>1,200,000</td>
</tr>
<tr>
<td>0.70 EUR/km</td>
<td>0.96 PPP/km</td>
<td>50,000</td>
<td>900,000</td>
<td>2 min</td>
</tr>
<tr>
<td>This paper</td>
<td>0.50 CHF/km</td>
<td>0.44 PPP/km</td>
<td>2,000</td>
<td>100,000</td>
</tr>
<tr>
<td>Zurich (Dynamic)</td>
<td>0.75 CHF/km</td>
<td>0.65 PPP/km</td>
<td>5,000</td>
<td>125,000</td>
</tr>
<tr>
<td></td>
<td>1.40 CHF/km</td>
<td>1.22 PPP/km</td>
<td>8,000</td>
<td>100,000</td>
</tr>
</tbody>
</table>

a Prices converted to USD in Purchasing Power Parity (OECD, 2020) in 2019 for Switzerland and France, and 2018 for Singapore.
b Approximative values either derived from the text or by examining the provided visualizations.
c Percentage of total demand defined in the study.
d Percentage of base price used in the study.
made on the operator profit or loss for these fleet simulations, our results suggest that for an average waiting time of three minutes most of the chosen prices are feasible when compared to the Paris case, but would be too low to have a cost-covering or profitable service in Zurich. Clearly, this also depends on other factors such as the density of the use cases, and especially on the potential user group that is attracted by the service. While all other studies in Table 3 analyse the demand for car (and, partly, public transport) users, our service is accessible to all travellers.

The striking difference between the upper and lower part of Table 3 is that in our cost-covering simulations higher prices are correlated with higher fleet sizes due to the simulation logic. Note that this would still be the case if we were to impose a profit margin on the price. Therefore, and because AMoD simulation studies tend to be geared towards very specific contexts, direct comparisons are difficult. Yet, results and assumptions on waiting times and prices across different methodologies seem to be consistent.

This is furthermore interesting as different fleet control strategies are used. While Hörl et al. (2019d) or Hyland and Mahmassani (2018) clearly show that the choice of the strategy has a strong influence on waiting times and driven distance, it is reasonable to think that their influence is diminished in studies where fleet size is either adapted to a fixed waiting time criterion or bounded by the acceptance of the travellers. Compared to other studies (Jing et al., 2020) the service simulated in this paper exhibits a rather high share of empty distance of almost 40% in some cases, while the common level of about 20% is only reached at around 6,000 vehicles. Interestingly, this case also falls closer to the range of waiting times and prices that have been imposed in other studies (Table 3).

The control strategy chosen here does not make use of re-balancing or pooling passengers. In a previous study (Hörl et al., 2019d), we have compared various operating strategies with static demand in Zurich. Despite its simplicity, the strategy used here provides a good trade-off between minimization of empty distance and minimization of waiting times. While certain algorithms (e.g. Pavone et al., 2012) can generally improve the service level (at static demand), we have shown that this comes at the cost of a comparably higher empty distance, especially for large fleet sizes. Therefore, it would be interesting to test various strategies with the framework presented here. Given that certain strategies perform best for lower or higher demand, their interplay with customer behaviour and acceptance is potentially highly complex and worth investigating.

The same is true for pooling. Various studies (e.g. Hyland and Mahmassani, 2020) show that it can increase the performance of a fleet if an algorithm dynamically aggregates demand. We think that the introduction of ride-pooling will have less ambiguous effects than the use of advanced (re-balancing) control strategies. Adding ride-pooling would effectively reduce empty distance and its associated cost. Consequently, waiting times would decrease at first, but the trips induced by cheaper prices would reoccupy the freed capacity. Therefore, if benefits of ride-pooling are not converted into profit (which is another option), it is a means of increasing equal access to the system if the cost-covering constraint is kept. Simulations including pooling are possible in our framework, and the survey data for Zurich even allows estimation of ride-pooling specific attributes such as the expected number of other customers in the vehicle. Yet, technical challenges remain which will be discussed later.

5.2. Implications for Zurich

In the following, the obtained results shall be further embedded in the context of Zurich. The city pro-actively fosters the use of bicycles while already a large share of travellers is using the dense and highly reliable public transport system. Mode shares (by distance) for public transport and active modes in the chosen operating area of AMoD today are at 44% and 38%, respectively. A relatively small share of 18% is covered by car (see Fig. 11). If AMoD is promoted as a tool to reduce and aggregate the use of individual traffic (as is often the case for studies on American cities), there is already little traffic to start with in the urban area of Zurich.

To put the obtained distance fares (Fig. 5) into perspective, they can be compared to the full cost of owning and using a private car in Switzerland, which is around 0.70 CHF/km. In the long term, an AMoD fleet of up to 4,000 vehicles would be competitive at that price level. While the relatively low increases in congestion (Fig. 13) indicate that road travel times would stay rather stable, the additional waiting time for the car becomes the crucial influence for, e.g., commute trips. As shown in Fig. 7 these waiting times can get large at times, especially for the evening commute.

In terms of pricing, it is interesting to look at the paid prices during simulation (Fig. 8) in comparison to public transport tickets. A single-ride ticket in Zurich costs 4.40 CHF, while a monthly subscription costs 85 CHF. The average distance travelled in the operating area is about 8 km, leading to a cost of 0.55 CHF/km for the single ride, and, assuming at least 50 rides per month for commuting, to 0.21 CHF/km for the subscription. In that regard, the AMoD service in all presented configurations is competitive price-wise for occasional use, but not in the long run. Again, the waiting time is the decisive factor, especially in remote areas away from the main corridors. Issues of equity can then arise when the AMoD systems draws paying customers from the public transport system. Taking into account the current level of subsidies of about 50%, a further decline in demand may threaten the access to well-functioning transport in areas that are today still densely covered with affordable public transport.

It is therefore important to think how an AMoD system could interact and potentially benefit the existing system. With increases of driven distance of almost 100% (Fig. 12) and mode shifts up to 20% (Fig. 11) the systems presented in this paper clearly do not fit into Zurich’s strategy of promoting active mobility and maintaining a high level of public transport use. A potential policy instrument for limiting the shift could be the introduction of a base fare for the service. Our experiments show (Fig. 10) that an increasing price indeed suppresses demand for the AMoD system from trips previously performed with active modes. Yet, the tested prices do not strongly affect the public transport share, which is mainly diminished at distances between two and four kilometers (Fig. 10). Arguably, an even higher base fare may have an effect and should be tested in future studies.

More importantly, the current simulations do not support inter-modal trips where travellers could use the AMoD system to access public transport infrastructure. Substantial research exists for various use cases and degrees of interaction (Shen et al., 2018; Salazar et al., 2018; Pinto et al., 2020; Gurumurthy et al., 2020b). The potential lies in the fact that current car users may give up their car if
they can reliably use an on-demand service to access the public transport system and have flawless interaction. Providing guarantees for arrival times and interchanges in an on-demand service is a topic of ongoing research in itself (Liu et al., 2019b). Technically, it is possible in our simulation and should be tested in the future. However, for the time being it is reasonable to argue that such interaction with the transport system (if not enforced) will not have a strong impact on the results as the chosen operating area is relatively dense. Taking an AMoD ride with several minutes of waiting time included would often exceed the time of walking to, and waiting for, a transit vehicle. However, the situation may change if remote areas are considered (Sieber et al., 2020). For Zurich, a natural and interesting use case would, therefore, be to study how an AMoD system might serve the commuter rail system connecting surrounding areas to the city center.

Interestingly, the enormous increase of driven distance on the roads (Fig. 12) does not lead to substantial increases in travel time (Fig. 13). Yet, it poses a large problem, given that the Federal Statistical Office (BFS) estimates an infrastructure maintenance cost of around 0.06 CHF/km (BFS, 2019). Having an AMoD operator (public or private) cover this cost would make sense. In our framework, this could be done easily by increasing the respective cost in Eq. 1 and should be done in a future study. The visualizations in Fig. 14 and analysis in Table 2 show that road usage merely increases in the sub-ordinate road network and not on large roads with are designed for high throughput and durability. The costs estimated by BFS may be therefore strongly underestimated in this case. Clearly, more research on the attractiveness of automated taxi services on short distance rides is necessary.

In this paper, we do not show any analyses of external costs of the automated taxi system. An analysis of GHG emissions would be interesting, as it may add another perspective to the assessment of the system. Since we already assume battery-electric automated vehicles a positive picture in terms of environmental impact may arise, given the already decreased number of vehicles (Fig. 12) and potential low emission electricity sources in Switzerland. Looking at the maps in Fig. 12, however, we expect a strong increase in cold emissions such as the noise produced and accidents caused in areas designed for low traffic volumes. The measurements of emissions, noise, congestion and accidents is possible in MATSim and studies of internalizing these costs have been performed (Kaddoura et al., 2020).

One effective policy for minimizing these externalities, which are mainly tied to the use of residential road infrastructure, may be to forbid the AMoD to enter such areas. This is clearly an interesting scenario to assess in future studies. However, it should be pointed out that the saved distance may be offset (at least partly) with maintenance trips, which are not considered in our simulation to date. Currently, we do not include important operational aspects such as recharging of the fleet and placement of charging infrastructure (Vosooghi et al., 2020; Zhang et al., 2020). Equally, vehicles pick up and drop off customers directly at the origin and destination of their trip (door-to-door). While this can be a policy-related question as discussed, it may also be operational as vehicles can not idle anywhere in the network at any time, especially if space is already occupied by other vehicles. So far, we have not performed analyses on how many vehicles reside at the same link at the same time in our network. Considering capacities for pick-up and drop-off points (PUDOs) is therefore an important component that needs to be added in future versions of the framework. Furthermore, not every place in the city is suited for idling for longer periods of time. The impact of parking search, restrictions, and strategies is therefore under active research (Harper et al., 2018; Bischoff et al., 2019; Wang et al., 2019; Winter et al., 2020; Yan et al., 2020), but neglected in this study. All of these components have the potential to increase the driven distance even more.

These considerations furthermore put the otherwise positive result of decreased vehicle use (Fig. 12) in question. While parking space may be freed up, our results suggest that this space would be filled up by taxis which are waiting or interacting with customers. Crossing the street in front of an automated vehicle may become a complex endeavor in itself (Rad et al., 2020); and overseeing potentially unpredictable behaviour of arriving and leaving vehicles may even increase this complexity.

5.3. Methodology

Methodologically, this paper shows how the agent- and activity-based transport simulation framework MATSim can be run with a survey-based discrete choice model to derive agents’ mode choice behavior. As described initially, many assumptions and concepts at the core of MATSim were discarded for this case. For instance, no scoring is taking place, where agents would introduce random changes to their daily activity chains, then test them in simulation and choose among in the past observed activity chains providing the best performance during simulation. Here, mode choices are imposed a priori, before simulating, but based on estimated choice dimensions from previous iterations. Through that, the choice process becomes more complex, but consistent with established discrete choice theory. In this context however, no other choices other than mode choice are considered. By default, MATSim also allows agents to update departure times and even their activities’ locations. As these processes are inherently based on the scoring approach, they can not be easily recovered in the present set-up. Promising new approaches to further integrate discrete choice models and the MATSim scoring approach are currently being researched.

Temporal consistency of the activity chains was only checked on a case-by-case basis in this research. As the choice model does not contain penalties for “being late” at the following activity in an agents’ chain, there is no guarantee that agents adhere to hard constraints such as latest arrival times at the job. However, such restrictions are also not modelled in the baseline population, which makes it hard to verify whether they are fulfilled. Looking at the distribution of departures by time of day (e.g. Fig. 7), we conclude that realistic patterns are established.

As one of the first studies in the MATSim ecosystem, we present a study with clear convergence criteria. While previous papers have stopped the simulation after a predefined number of iterations, we stop the simulation when all endogenous key metrics have converged. Clearly, our convergence criteria involve trade-offs as well. Defining them more narrowly would further stabilize the simulation results, but would also drive up simulation time. In comparison to earlier work, however, we avoid stopping the simulation at an arbitrary state as is shown by our analysis of different run times (and iterations until convergence) for different fleet sizes. While
the chosen criteria can certainly be improved, they represent a starting point for further developments in that direction. In another stream of research, we want to include standard chain convergence diagnostics into MATSim (Heidelberger and Welch, 1983; Geweke, 1992).

Run times still pose a large obstacle for our simulations. As outlined in Appendix B, one simulation takes about 14 h and 200 iterations for 8,000 vehicles, which makes about 4 min per iteration. In (Hörl et al., 2019d) we compare different dispatching strategies with the heuristic used in this paper. For a demand of 100% and a larger operating area around the center of Zurich, the heuristic simulation takes one hour to execute one iteration. More intelligent algorithms, like (Pavone et al., 2012), where a linear program is frequently solved to redistribute vehicles, are measured to take about four times longer. Applied to the present case this would mean that a full multi-iteration simulation until equilibrium could take up to 56 h. This is still feasible, but poses limits on the ensemble size that can be tested and the number of different cases. Yet, running the simulations presented here with more advanced algorithms (Ruch et al., 2020; Hörl et al., 2019d; Hyland and Mahmassani, 2017; Mourad et al., 2019) will be an important extension in future research.

This includes the use of pooling algorithms, e.g. by Liang et al. (2020) or the high-capacity dispatcher by Alonso-Mora et al. (2017), which is already implemented in the AMoDeus framework3 (Ruch et al., 2018), a collection of fleet algorithm implementations that is compatible with our model set-up. For better simulation performance, the latter algorithm has been augmented with intelligent heuristics and applied in recent research (Bilali et al., 2020).

The use of pooling algorithms will likely make it necessary to switch to a larger sample size of the population than 10%. While it may not be necessary to run all future simulations with 100% of the demand, it will be necessary to systematically analyze at which level of down-sampling results do not differ too much from the ideal case.

Finally, we want to comment on the replicability of the study. While the authors have presented methods and tools to produce a detailed agent-based transport demand for Ile-de-France and Paris based entirely on open data and open software (Hörl and Balac, 2020), this is not the case for Switzerland; the data sets involved are only provided for research, without sufficient anonymization. For instance, we generate the underlying population directly from the Swiss census data set, which contains the location by coordinate for every household, including the socio-demographic attributes of its members. In the future, it will be important to focus on generative algorithms that can replicate the Swiss population without conflicting with privacy concerns and to further engage with the responsible authorities to make those data sets easily and openly accessible. The code used to run the simulations in the present paper is available as part of the eqasim framework4, which aims to consolidate tools and data to run standardized, open, high-quality, and out-of-the-box runnable MATSim simulations.

6. Conclusion

To conclude, it can be stated that the autonomous mobility on-demand system configurations presented in this paper clearly do not support a road-map towards a cleaner, widely accessible, more active and livable city. The major insights are therefore: (1) A cost-covering, but otherwise widely unregulated AMoD operator would reduce private car use, but at the same time cause large infrastructure costs, increases in noise, congestion and accidents. (2) A commercially operated service would likely lead to fewer trips and less congestion, but would still draw demand from active modes and public transport at rather expensive fares for those who can afford them. (3) Applying further regulation on any form of AMoD system in Zurich will be necessary.

Relevant policy measures include:

- **Spatial restrictions**: An AMoD system should only be allowed to pick up and drop off customers in designated areas (PUDOds) or at major road infrastructure.
- **Pricing**: To avoid modal shifts from active modes and public transport, a minimum price should be introduced where high competition with public transport is expected. To accommodate the high expected deterioration of road infrastructure, mandatory reporting and taxation of driven distances may be required.
- **Inter-modality**: When designing a publicly operated system, focus should be put on its inter-operability with the public transport system. Areas of high public transport density may only allow first- and last-mile traffic.
- **Sharing**: Incentivising ride-pooling may increase the equal access to the system.

These policy recommendations are based on insights derived from our results, which show that an AMoD system in Zurich can bring benefits to the users, but has largely negative impacts on the transport system. The recommended policies have not been validated in simulation, which should be done in a follow-up study. Furthermore, the simulation framework should be extended in the future with additional operational aspects, more intelligent fleet control algorithms and ride-pooling.

CRediT authorship contribution statement

**Sebastian Hörl**: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration. **Felix Becker**: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Project administration. **Kay W.**

---

Axhausen: Conceptualization, Resources, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgements

This paper is based on research that was commissioned by the Association of Swiss Transport Engineers and Experts (SVI) and funded by the Federal Road Administration of Switzerland in the scope of the project SVI 2016/001: Induced demand by automated vehicles (Hörl et al., 2019c). We want to thank Karen Ettlin for editing an earlier version of the manuscript.

Appendix A. Baseline model

The following sections describe how our baseline simulation for Zurich was set up. The first section describes which data sets were used and processed, while the second section covers the calibration and validation with reference data of the model.

A.1. Data sources and processing

The basis of an agent-based MATSim simulation is a synthetic population of the case study area. For Switzerland, a range of rich data sources are available that make it possible to set up such a population with detailed socio-demographic and behavioral attributes. In the following, the process will be described in brief. A more detailed description of the Swiss data sets and how they have been processed can be found in (Hörl, 2020). Furthermore, (Hörl and Balac, 2020) describes the methodology in detail for the case of Paris and Ile-de-France. There, unlike Switzerland, all necessary data sets are publicly and openly available to replicate the process.

Initially, the synthetic Swiss population is based on the STATPOP data set, which is a comprehensive census of Switzerland, including attributes like age and gender of all residents, but - most notably - also location by coordinate of each household. Persons and households from this data set are used directly as a basis of the Swiss synthetic population. Afterward, work or educational locations for each synthetic person are added. For that, the Structural Survey is used, which, unlike the registration-based census data, is a detailed household survey conducted for around 3% of the Swiss population each year. From this data set, an origin–destination matrix between home and work/education municipalities in Switzerland can be derived, then used to proportionately sample destination zones for each agent, based on their home zone. In the third step, the national Mobility and Transport Microcensus is employed: a classical household travel survey providing full day activity patterns for around 60,000 respondents. We use a statistical matching procedure to sample and attach activity chains to the synthetic persons, based on socio-demographic attributes. This way, activity patterns correlating with certain age groups or gender are assigned according to the population. These activity chains contain activity types, activity durations and connecting modes of transport as reported in the survey data. As departure and arrival times are rounded strongly (in five minute intervals) as is the case in many surveys, activity chains (including all departure and arrival times) are shifted by a random offset of +/- 30 min for each person to avoid having unrealistically many people departing at the same time at one location, which would, in turn, affect the traffic dynamics in the simulation negatively. Finally, specific work and education locations are sampled, based on the destination zones of the agents, from STATENT, the national enterprise census, which lists all facilities in Switzerland by coordinate. Further, secondary activities in the agents’ plans (such as shopping or leisure), are assigned to discrete facilities using a specifically-designed algorithm for secondary location assignment (Hörl and Axhausen, 2020). To complete the simulation data, an exhaustive digital public transit schedule for Switzerland is integrated into the simulation and the road network is extracted from OpenStreetMap, including all roads from motorways to residential roads. In total, this process yields a synthetic representation of Switzerland that reproduces the daily mobility needs of the population and available infrastructure.

A.2. Calibration

This data serves as an input to a MATSim simulation. Here, it is used in combination with a new Discrete Mode Choice package presented by Hörl et al. (2019a). Such a simulation runs multiple iterations between mobility simulation and the choice process. First, initial synthetic population plans are simulated in the capacitated network. As the synthetic population contains predefined transport modes from the household travel survey, some roads are unrealistically empty or congested. Travel times on the links are tracked and then, a small percentage (here 5%) of agents performs new mode choices for all trips in their daily plan based on the model described above, but also based on measured travel times. Hence, car trips with heavy congestion are avoided. In the next iteration, different agents use their car at different times during the day, leading to new travel times, which are, in turn, fed back to the next round of decision making. This way, we simulate a large-scale mode choice experiment with a synthetic population. Note that MATSim, in principle, is also able to consider departure time choices or even location choice through its co-evolutionary algorithm, which is a different process of decision-making not used here; our goal was to directly integrate a discrete choice model into MATSim. Unlike the standard approach, it inherently contains consistent information about the trade-offs people make between choice dimensions when faced with the automated mobility option. The standard co-evolutionary algorithm would only allow us to perform sensitivity analyses on the relatively unspecified attractively level of the new transport mode.

For this study, a perimeter of 30 km around the city center of Zurich is cut out from the Switzerland model and population. Through-traffic, as well as agents who enter or exit the area at some point during the day, are included. In total, this gives around 2.2 million agents performing 7.2 million activities and a network of around 150,000 links.

Due to the complexity of the simulation, the dense network and the large number of agents, we restrict simulations to 10% samples of the population, with road capacities being scaled down accordingly. These capacities need to be scaled in any case, as we do not
model explicitly traffic from tourists, deliveries or heavy freight.

As it is difficult to down-scale the public transit supply (in terms of frequency or vehicle capacity), and to further reduce run time, we simulate public transport deterministically according to schedule. This means that agents look up the fastest connection from an origin to a destination (at a specific departure time) from the transit schedule using the RAPTOR algorithm (Delling et al., 2015), and they appear at their destination at the planned arrival time. Clearly, this approach does not take into account perturbations or delays caused by traffic. However, as the public transit system in Zurich is rather stable and delays are rare, we consider this as a minor assumption.

The simulation was run with the choice model described above for calibration, without the AMoD mode included. In our simulation set-up, the trip-based model is served with the respective trip attributes, but choices are made on the level of home-based tours. For that, all possible combinations of modes along a tour are constructed. Then, for each chain of modes, the trips’ cumulative utility is calculated. Based on these cumulative utilities, one chain is chosen according to the logit formula. In combination with tour-level constraints (e.g., a car or a bike can not be left behind, but must be brought back home), this approach has been shown to provide realistic mode shares (Hörl et al., 2019a; Hörl et al., 2019b). However, due to specific survey characteristics, such as generally long distances asked from the respondents, adjustments needed to be added to the model, especially to achieve a good fit with reference mode shares for small distances.

To correct for an increased attractiveness of the walk mode, its utility function is adjusted as follows:

\[ u'_{walk} = u_{walk} - \exp \left( \log(101) \cdot \frac{x_{travelTime, walk}}{\theta_{walkThreshold}} \right) + 1 \]  

(A.1)

For small walk distances, the additional penalty term is close to zero, while for a walking time equal to the threshold calibration parameter \( \theta_{walkThreshold} = 120 \text{ min} \), there is a large offset of \(-100\).

Furthermore, the value of \( \beta_{ASC, car} \) needed to be adjusted to achieve a good model fit. The value was changed from \( 0.223 \) to \( \beta_{ASC, car} = -0.8 \). This offset equals a perceived cost of around 11CHF and can be explained by a couple of factors, which would be interesting to include in future surveys and models. Among them are the additional cost of parking for car trips, a general perception of sunk costs (not included in the per-distance cost), time spent for parking search and the access/egress time.

Fig. A.15 shows mode shares in the baseline model compared to reference data from the household travel survey. Note that both
simulation data and reference data are filtered for trips that take place entirely within the study area. It can be seen that the reference data is rather sparse e.g. for bike, and other modes for longer trips), but that the simulated mode shares closely follow the curve shapes. The plots also show how the adjustments described above have improved the model fit. Making sure that the mode choice behavior is consistent with reality was our primary calibration objective, because changes towards an AMoD system can only be meaningful in comparison to a solid baseline case. It should be noted that the "car passenger" mode was not considered explicitly in the survey, but needs to be included in the simulation to align with a consistent relation between agents’ car ownership attribute and the modal split for private cars. Therefore, trips with this mode are kept fixed as generated from the population synthesis step, based on the household survey data.

The second calibration objective was to achieve travel times that fit reality well. This can be seen in Fig. A.16, where the mean, median and 10% and 90% quantiles of car trip durations are shown in distance classes of 500 meters, both for the uncalibrated and calibrated simulation and for the household travel survey. Note that these are average travel times calculated by the difference between the time of departure and arrival. Achieving a good fit in terms of the travel time distribution and still keeping a good fit with mode shares required to adjust the flow capacities of the generated road network simultaneously with the parameters mentioned above. The standard values that are defined by the pt2matsim\(^5\) converter and based on OpenStreetMap road categories and the Highway Capacity Manual were used and scaled down uniformly.

Appendix B. Convergence and run time

The simulation feedback loop in our set-up has three major components. The decisions of the choice model depend on travel times in the network (which are used to calculate the expected travel time for a trip with the car or the automated taxi); the waiting times achieved in the last iteration, measured by zones and time bins; and the distance fare of the AMoD, calculated on the basis of measured number of requests, driven distance with and without customers, and the fleet size. The major task of the simulation is the number of requests attracted by the AMoD system. The model is set-up in a circular fashion with gradual adaptation of the mode choices to the given travel times, waiting times and prices to reach an equilibrium state.

Decisions, although based on the discrete choice model and a structured utility function, include a certain amount of randomness. Especially, the measurement of waiting times shows noisy behaviour in our experiments. While waiting times in one zone at a certain time may be low in one iteration, this may lead to many agents choosing to use the service in that bin in the next iteration. The generated demand may then be too high for the fleet to serve and high waiting times arise in that zone. While this would lead to strong fluctuations in an all-or-nothing assignment of transport modes, only a fraction of the population (5% per iteration) performs mode choices. This process dampens the process where otherwise strong oscillations could happen.

To decide when a simulation is finished, we define two convergence criteria, which shall be explained with the number of requests as the major metric. We measure the number of requests for the AMoD service \(x_i \in \mathbb{N}\) in every iteration. As the demand starts at zero, but we do not know which final value will be reached in equilibrium, we can not simply measure the distance of this value to a known target. Instead, the idea is to measure the difference of the instantaneous value \(x_i\) in comparison to the mean that has been established over the horizon \(H \in \mathbb{N}\) of past iterations. The mean in iteration \(i\) with horizon \(H\) can be defined as

\[
\mu_i = \begin{cases} 
\frac{1}{H} \sum_{k=1}^{H} x_{i-k} & \text{if } i \geq H \\
\infty & \text{else}
\end{cases}
\]  

(B.1)

If not enough iterations have been simulated to establish the mean over horizon \(H\), an infinite value is assumed. The first convergence criterion is then to test whether the instantaneous values is within a certain threshold \(T_\mu\) from the mean:

\[|x_i - \mu_i| < T_\mu\]
This threshold makes sure that in the converged phase of the simulation the selected iteration does not vary too much from the established mean and, therefore, is not a strong outlier which may normalize again over the following iterations.

As indicated by the fact that demand starts from zero, there is a strong transient phase in the $x_i$ signal, which means that the average itself is changing strongly in the beginning of the simulation. We, therefore, define a second criterion to make sure that the simulation is in equilibrium state by comparing the current mean with the mean that was measured at some iteration before, described by the lag $L$. The lagged mean is defined as:

$$|x_i - \mu_i| \leq T_{\lambda}$$  
(B.2)

Table B.4

Convergence parameters.

<table>
<thead>
<tr>
<th>Metric</th>
<th>$T_{\mu}$</th>
<th>$T_{\lambda}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requests$^a$</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>Distance fare [CHF/km]</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Waiting time error [s]</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>Travel time error [s]</td>
<td>60</td>
<td>1</td>
</tr>
</tbody>
</table>

$^a$ Upscaled values of the 10% simulation to be in line with graphs in the main text.
\[ \lambda_i = \begin{cases} 
\mu_i - L & \text{if } i > H + L \\
\infty & \text{else}
\end{cases} \]  

(B.3)

As before, a criterion can be defined to require that the current mean is within a certain threshold \( T_y \) of the lagged mean, i.e.

\[ |\lambda_i - \mu| < T_y \]  

(B.4)

Both constraints, for the instantaneous value of the quantity in comparison to the mean, and the one for the comparison of the mean to the lagged mean, are applied to the four metrics number of requests, distance fare, waiting time error, and travel time error.

The latter two are defined by first temporarily saving the estimated values of travel and waiting time of a certain trip during the choice process. If the trip is chosen and simulated, we then compare the estimated values with those observed in the simulation. The “error” is then defined as the difference between estimated and simulated value. A positive error therefore corresponds to an underestimation of the value, while a negative value corresponds to overestimation. To arrive at a metric that can be used within the convergence strategy above, we calculate the mean error over all trips performed during the iteration. It is then possible to require with the convergence criteria that this average error stabilizes around a mean value over the course of multiple iterations and that the current iteration stays within close distance to this mean.

The choice of the correct values of horizon \( H \), lag \( L \) and the thresholds \( T_y \) and \( T_r \) required some preliminary simulations with many iterations on which various combinations of values were tested. The choice of the convergence parameters needs to make a trade-off between acceptable run times and acceptable stability of the obtained equilibrium. The analyses lead to the use of \( H = 50 \) and \( L = 25 \) for all four metrics in our simulations. The specific thresholds are specified per metric as defined in Table B.4.

We hence require that the difference between the two means not be larger than 1,000 requests (while the magnitude of the final value is between 50,000 and 100,000) and that the final value is within this range from the mean. For the distance fare we impose strict restrictions that the differences should not exceed 0.01 CHF. For waiting and travel times we require the instantaneous value to be within 15 s and 60 s from the respective mean. The two means we require to not vary more than one second from the lagged means.

Fig. B.17 shows example simulations with three different fleet sizes. The rows show the four relevant metrics for convergence. The simulation stops once all eight (2 criteria x 4 metrics) conditions are fulfilled. Note that the number of requests and the active price have rather smooth trajectories, so the \( T_y \) condition is not as important as for the travel and waiting time error metrics, which are rather noisy.

The plots show the metric value in each iteration, the mean over horizon \( H \) and the lagged mean. Only when these values get sufficiently close (according to Table B.4) the condition for convergence are met. There is a general pattern that simulations with larger fleet sizes need longer to converge. The first influence on convergence time is the final number of requests. As mode choice happens for only 5% of randomly selected agents in each iteration, demand is building up slowly and it takes longer until all demand is discovered. However, the difference in demand between fleet sizes 4,000 and 6,000 is not large in the shown examples. One major influence of simulation length is therefore the chosen way of defining the cost. Note that initially, the distance fare of 0.7 CHF is used for both 4,000 and 6,000 vehicles. However, for 4,000 vehicles the system switches to the calculated price of about 0.9 CHF after 25 iterations, while it switches to almost 1.2 CHF with 6,000 vehicles. It then takes longer for the price model to go into equilibrium with the demand.

It is interesting to note that the system diverges for fleet sizes larger than around 10,000 vehicles. In those cases the jump is large enough such that in the following iteration demand directly declines. Based on lower demand, prices increase in the following iteration such that even more demand is lost. Increasingly fast the distance fare then tends towards infinity while the demand falls down to zero over the next iterations.

Fig. B.17 also shows that travel times are generally underestimated, though less than 10 s on average. Predicting travel times with higher accuracy is difficult with the dynamic queuing model that is used in MATSim as complex interactions of demand and capacity can arise. Also the simulation runs on one-second time steps while average travel times per road link and time bin can be measured as continuous values. The choice between rounding up or down the measured travel times can then make a difference for routes that traverse many different links.

One iteration of the simulation takes around 220 s on modern hardware, leading to run times of around 8 h for small fleet sizes to 14 h for large ones. In total all simulations run in this paper (with the different fleet sizes, 3 base fares per fleet sizes, and 5 random seeds per combination) amount to a total run time of 1,627 h in computation time. Note that in (Hörl et al., 2019d) we present simulations very similar to the one in this paper in terms of perimeter, but without the decision-making loop of the travellers. There, we simulate a population that is not down-sampled with about 1 h per iteration. Translated to the present case we would expect run times of about 80 h to 140 h, which is still in an acceptable range if parallel computation infrastructure is available.

References


