Technische Universität München
TUM School of Engineering and Design

Operation and Regulation of Autonomous Mobility-on-Demand Systems

Florian Dandl

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung des akademischen Grades eines *Doktors der Ingenieurwissenschaften (Dr.-Ing.)* genehmigten Dissertation.

Vorsitz: Prof. Dr. Constantinos Antoniou
Prüfer*innen der Dissertation:
1. Prof. Dr.-Ing. Klaus Bogenberger
2. Prof. Dr. Hani S. Mahmassani
3. Prof. Dr. Maximilian Schiffer

Die Dissertation wurde am 25.01.2022 bei der Technischen Universität München eingereicht und durch die *TUM School of Engineering and Design* am 17.08.2022 angenommen.

Acknowledgment

I would like to begin by thanking my PhD supervisor, Prof. Klaus Bogenberger, for giving me the opportunity to work at this chair and become a PhD student. I highly appreciate the discussions about research ideas, his way of communicating his opinions, and his advice to help me move forward. His interest in the personalities of his PhD students and group dynamics are certainly one of the reasons why the atmosphere working at his chair has been a pleasure. Additionally, our countless private conversations helped to build enormous trust.

Next, I want to thank Prof. Hani S. Mahmassani for agreeing to be my second supervisor. After a very friendly getting to know each other at TRB, he offered great advice and feedback from very detailed questions for one research paper to the general research direction for this thesis. It has been an honor to work with this distinguished researcher, who has both a very fond knowledge of existing research and visions for the future.

My gratitude also goes out to Prof. Michael F. Hyland. My academic career would not have been the same without him. Mike's friendly and calm nature, combined with his knowledge and skill made it a great pleasure to collaborate. I greatly benefited from our discussions on all levels, from ideas and methodological approaches down to single lines of code and from the overall interpretation down to single phrases.

I have been working on some interesting projects that allowed me to gain insights into mobility services and create a network in industry and academia. I want to thank the project partners at MOIA, BMW, SWM, LHM, and the researchers at KIT-IfV and other TUM chairs for the welcoming and productive collaborations.

I want to express my gratitude towards all of my former and current coworkers — at UniBW and then at TUM. Even though I spent a lot of time traveling to and from work, it was worthwhile to go to the office thanks to the great atmosphere. I want to point out Dr. Antonios Tsakarestos for taking the role as mentor, and Gabriel Tilg, Dr. Majid Rostami, Fabian Fehn, and Patrick Malcolm for our collaborations. Although Aledia Bilali, Marvin Erdmann, and Arslan-Ali Syed were external PhD students, I regarded them as colleagues due to their effort to integrate with the chair. It was a pleasure to develop ideas and research papers together.

Moreover, I want to highlight a few of my fellow UniBW (& TUM) PhD students: Lisa Kessler, who somehow manages to get a lot of work done and have a private life; Cornelius Hardt, whose critical view often helped to improve a situation; Dr. Benedikt Bracher, who never missed an opportunity to cheer everybody up; Dr. Gerhard Huber, who has been my friend long before starting my PhD and made me aware of this opportunity; Tanja Niels, my prior roommate who was ready to discuss any problem without preparation; and lastly, but most importantly, Roman Engelhardt. Somehow, the two of us just understood each other and the way we think with the day he started. Since then, we have worked on several projects together and developed a simulation framework for the chair which is used in a multitude of projects.

Finally, I would like to thank my friends and family for their support. The effort I was able to put into my work was only possible because I could fall back on my private life. I am exceptionally thankful to my parents Monika and Georg Dandl for their unending love and support. But, most of all, I am very grateful to my wife, Diana Dandl. Her love is the foundation I build my life on. I cannot thank her enough.

Abstract

Technological advancement, more specifically digitalization, mobile internet, and the widespread availability of smartphones enabled the rise of mobility-on-demand services. In carsharing systems, users can rent a vehicle, walk to, and drive with it for one-way or round trips. Ride-hailing services do not even require the users to drive anymore. After booking, a driver picks up and drops off the user at the user's location and the desired destination, respectively. Carsharing and ride-hailing offer a similar comfort level as a private vehicle, but without some of the downsides of owning a private vehicle. For instance, the fixed costs of cars are shared among multiple users, and the search for a full-time parking slot is either simplified (carsharing) or not necessary (ride-hailing). Due to its high convenience, user-centric business model, and comparatively affordable prices, ride-hailing services have become more popular than taxis, grown tremendously, and are a relevant part of urban transportation systems in many cities nowadays. As a downside, the extensive use of ride-hailing has led to increased levels of total vehicle kilometers and congestion in some cities. Ride-pooling, which tries to match multiple users into a single vehicle, can help to mitigate these negative externalities.

Another technological advancement on the horizon is self-driving vehicle technology. In today's ride-hailing services, the drivers are the highest cost component. With a much cheaper cost structure, the mobility service can be offered at a much cheaper price point, probably attracting a significant amount of demand. From a user perspective, an autonomous mobility-on-demand service will likely look very similar to today's ride-hailing service, except for the vehicle being without a driver. Initially, this could lead to a different experience of the service, but — as with other new technological advancements — become the new normal over a short period of time. From an operator's perspective, some tasks are different than today. In most of today's ride-hailing services, the drivers are ultimately the decision-makers. They decide when to start and end their service hours, which customers to accept, and when and where to cruise in the city. With autonomous vehicles, each vehicle can be controlled directly by the operator and — besides maintenance — the availability can be 24/7. Therefore, the potential for new service optimization techniques emerged.

The first major research question asks how to operate the fleet centrally: What are the most important tasks/problems and how to address them? An operator of an autonomous mobility-on-demand service has to make several decisions regarding service design parameters, the interaction with users, and vehicle routing. For example, an service provider has to specify whether to offer a pooling service, the general fare, and time constraints for waiting and detour time. Moreover, operational aspects have to be addressed: fleet size, repositioning vehicles to be close to possible future requests, estimating the level-of-service and dynamically setting the fare to make offers to requests, and assigning these requests to vehicle tours in case of a booking. The combination of these tasks can be viewed as a rich dynamic and stochastic vehicle routing problem. The thesis contributes in this part by the development of a new solution approach to this difficult-to-solve vehicle routing problem. Decisions or actions have impacts on different time scales. For instance, an assignment of a request to a vehicle affects the route of this vehicle for a few minutes. Contrarily, the set of all repositioning vehicles should equalize the balance of demand and supply over a much longer period of time. Selecting a fleet size can even be considered a planning process, which is only performed very

infrequently as the acquisition of vehicles takes some time and once the investment is made, vehicles should be utilized. Based on the different time scales, the large operational problem is divided into more manageable sub-problems, for which solution approaches are developed and evaluated in a case study. Insertion heuristics and global optimization are studied for the offer and vehicle-assignment processes, which are required reactions to user requests. With the help of forecasts, tactical decisions to balance supply and demand can be utilized, e.g., for repositioning and dynamic pricing. Typically, the operating area is divided into independent zones for the respective strategies. New approaches to repositioning and dynamic pricing, which are based on the inter-dependency of these zones, are developed in this thesis. Operators ultimately will invest computational resources into all of the sub-problems to maximize their profits. However, the long-term planning variables fleet size and pricing are identified as those decisions that will influence the demand and profit the most.

The second major research question takes the perspective of a regulating public entity and traffic planner: How will an autonomous mobility-on-demand service affect the urban transportation system and how to regulate this system? Compared to typical tasks of traffic planners, where a measure is developed and the change in user behavior often leads to a bi-level problem, the operators of mobility services should be considered as a decision-making entity as well. The contributions with respect to this research question are the development of a tri-level problem formulation, a solution approach with Bayesian optimization, and the creation of an agent-based transportation model, which has to trade-off model detail and computational effort. The tri-level model separates the time scales with the assumptions that regulators typically require a lot of time to plan and conduct their actions and regulations should not be changed on a daily basis. Operators have sufficient time to adapt their service to this regulatory setting to maximize their profit and offer a stable service. On the third level, travelers choose between different modes for their daily mobility needs, with the autonomous mobility on demand service being one of these modes. The multi-level Bayesian optimization model developed in this thesis helps to address this tri-level problem as the next step in this iterative optimization procedure can infer knowledge from prior iterations in this multi-level multi-dimensional solution space. Even with this solution approach, the transportation model has to be computationally very efficient. Therefore, the transportation model aggregates where possible but still keeps detailed models for the most critical aspects. An agent-based approach is chosen to model the decision-making of travelers in a tractable fashion. Moreover, the matching of users to vehicle routes — especially for pooling — is hardly assessable on a macroscopic scale. Additionally, network dynamics have to be modeled as the decisions of travelers should affect the transportation system.

In general, autonomous mobility-on-demand services add a new option for travelers that significantly increases the population's travel utility and can be viewed as a pull measure, which can be expected to cause significant modal shifts. The developed framework helps to identify how to set push measures like toll or parking fees. Furthermore, the sensitivity analysis of these measures allows noticing when they become too severe, i.e., the regulations reduce the utility of traveling without bringing social benefits by improving the transportation network as a whole.

In summary, this thesis contributes to a better understanding of how to operate autonomous mobility-on-demand systems and how public entities can deal with the new mobility system.

List of Publications

My general understanding of the topic have developed over the last years and lead to several publications. Therefore, some ideas in this dissertation have appeared previously in the publications listed below:

Journal Papers

1. SYED, A.A.; DANDL, F.; KALTENHAEUSER, B.; BOGENBERGER, K. (2021). "Density Based Distribution Model for Repositioning Strategies of Ride Hailing Services", Frontiers in Future Transportation, Volume 2, pp. 295

2. ERDMANN, M.; DANDL, F.; BOGENBERGER, K. (2021). "Combining Immediate Customer Responses and Car-Passenger Reassignments in On-Demand Mobility Services", Transportation Research Part C, Volume 126(10), pp. 103104

3. WILKES, G.; ENGELHARDT, R.; BRIEM, L.; DANDL, F.; VORTISCH, P.; BOGENBERGER, K.; KAGERBAUER, M. (2021). "Self-Regulating Demand and Supply Equilibrium in Joint Simulation of Travel Demand and a Ride-Pooling Service", Transportation Research Record: Journal of the Transportation Research Board, pp. 036119812199714

4. DANDL, F.; ENGELHARDT, R.; HYLAND, M.; TILG, G.; BOGENBERGER, K.; MAHMASSANI H. (2021). "Regulating Mobility-on-Demand Services: Tri-level Model and Bayesian Optimization Solution Approach", Transportation Research Part C: Emerging Technologies, Volume 125, p103075

5. BILALI, A.; ENGELHARDT, R.; DANDL, F.; FASTENRATH, U.; BOGENBERGER, K. (2020). "An Analytical and Agent-Based Model to Evaluate Ride Pooling Impact Factors", Transportation Research Record: Journal of the Transportation Research Board, pp. 036119812091766

6. KALTENHAEUSER, B.; WERDICH, K.; DANDL, F.; BOGENBERGER, K. (2020). "Market Development of Autonomous Driving in Germany", Transportation Research Part A: Policy and Practice, Volume 132 pp. 882–910

7. HYLAND, M.; DANDL, F.; BOGENBERGER, K.; MAHMASSANI, H. (2020). "Integrating Demand Forecasts into the Operational Strategies of Shared Automated Vehicle Mobility Services: Spatial Resolution Impacts", Transportation Letters, Volume 12

8. DANDL, F.; HYLAND, M.; BOGENBERGER, K.; MAHMASSANI, H. (2019). "Evaluating the Impact of Spatio-Temporal Demand Forecast Aggregation on the Operational Performance of Shared Autonomous Mobility Fleets", Transportation, Volume 114, pp. 462-484

9. DANDL, F.; BOGENBERGER, K. (2019). "Comparing Future Autonomous Electric Taxis With an Existing Free-Floating Carsharing System", IEEE Transactions on Intelligent Transportation Systems, Volume 20, Issue 6, pp. 2037-2047

Conference Publications (Full Paper with Review)

1. ENGELHARDT, R.; MALCOM, P.; DANDL, F.; BOGENBERGER, K. (2022) "Competition and Cooperation of Autonomous Ridepooling Services — Simulation of a Novel Broker Concept", Transportation Research Board 101st Annual Meeting, Washington, D.C., USA

2. HARDT, C.; DANDL, F.; BOGENBERGER, K. (2022) "Event-Driven Pricing in Mobility Sharing Systems", Transportation Research Board 101st Annual Meeting, Washington, D.C., USA

3. WILKES, G.; ENGELHARDT, R.; KOSTORZ, N.; DANDL, F.; VORTISCH, P.; BOGENBERGER, K.; KAGERBAUER, M. (2022) "Autonomous Ride-Pooling in Hamburg — Results from an Integrated Demand and Fleet Simulation Model", Transportation Research Board 101st Annual Meeting, Washington, D.C., USA

4. WILKES, G.; ENGELHARDT, R.; KOSTORZ, N.; DANDL, F.; ZWICK, F.; FRAEDRICH, E.; KAGERBAUER, M. (2021) "Assessing the Effects of Ridepooling — A Case Study of MOIA in Hamburg", 27th ITS World Congress, Hamburg, Germany

5. DANDL, F.; ENGELHARDT, R.; BOGENBERGER, K. (2021). "On the Dynamism of User Rejections in Mobility-on-Demand Systems", 24th IEEE International Conference on Intelligent Transportation Systems (ITSC), Indianapolis, USA

6. SYED, A.A.; DANDL, F.; BOGENBERGER, K. (2021). "User-Assignment Strategy Considering Future Imbalance Impacts for Ride Hailing", 24th IEEE International Conference on Intelligent Transportation Systems (ITSC), Indianapolis, USA

7. DANDL, F.; TILG, G.; ROSTAMI-SHARBABAKI, M.; BOGENBERGER, K. (2020). "Network Fundamental Diagram Based Routing of Vehicle Fleets in Dynamic Traffic Simulations", 23rd IEEE International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece

8. DANDL, F.; FEHN, F.; BOGENBERGER, K.; BUSCH, F. (2020). "Pre-Day Scheduling of Charging Processes in Mobility-on-Demand Systems Considering Electricity Price and Vehicle Utilization Forecasts", Forum on Integrated and Sustainable Transportation Systems (FISTS), Delft, Netherlands

9. DANDL, F.; NIELS, T.; BOGENBERGER, K. (2020). "Design and Control of Park & Charge Lanes for Carsharing Services with Highly-Automated Electric Vehicles", 21st IFAC World Congress, Berlin, Germany

10. DANDL, F.; HYLAND, M.; BOGENBERGER, K.; MAHMASSANI, H. (2020). "Dual-Horizon Forecasts and Repositioning Strategies for Operating Shared Autonomous Mobility Fleets", Transportation Research Board 99th Annual Meeting, Washington, D.C., USA

11. ERDMANN, M.; DANDL, F.; KALTENHAEUSER, B.; BOGENBERGER, K. (2020). "Dynamic Car-Passenger Matching of Online and Reservation Requests", Transportation Research Board 99th Annual Meeting, Washington, D.C., USA

12. BILALI, A.; DANDL, F.; FASTENRATH, U.; BOGENBERGER, K. (2019). "An Analytical Model for On-Demand Ride Sharing to Evaluate the Impact of Reservation, Detour and Maximum Waiting Time", 22nd IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand

13. ENGELHARDT, R.; DANDL, F.; BILALI, A.; BOGENBERGER, K. (2019). "Quantifying the Benefits of Autonomous On-Demand Ride-Pooling: A Simulation Study for Munich, Germany", 22nd Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand

14. DANDL, F.; BOGENBERGER, K.; MAHMASSANI, H. (2019). "Autonomous Mobility-on-Demand Real-Time Gaming Framework", 6th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Krakow (Poland)

15. BILALI, A.; DANDL, F.; FASTENRATH, U.; BOGENBERGER, K. (2019). "Impact of service quality factors on ride sharing in urban areas", 6th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Krakow (Poland)

16. ERDMANN, M.; DANDL, F.; BOGENBERGER, K. (2019). "Dynamic Car-Passenger Matching based on Tabu Search using Global Optimization with Time Windows", 8th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO), Bahrain

17. DANDL, F.; GRUEBER, B.; FRIESE, H.; BOGENBERGER, K. (2018). "Design and Simulation of a Public-Transportation-Complimentary Autonomous Commuter Shuttle", mobil.TUM 2018, Munich, Germany

18. DANDL, F.; BOGENBERGER, K. (2018). "Booking Processes in Autonomous Carsharing and Taxi Systems", 7th Transport Research Arena TRA 2018, 2018, Vienna, Austria

19. DANDL, F.; BRACHER; B.; BOGENBERGER, K. (2017). "Microsimulation of an Autonomous Taxi-System in Munich", 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Napoli, Italy

Contents

Chapter 1

Introduction

The *operation and regulation of autonomous mobility on demand (AMoD) systems* were selected as the topics of this thesis. Even though the path and timeline towards autonomous vehicles and AMoD systems are still uncertain, it is essential to analyze AMoD systems already today as they can potentially disrupt transportation systems. Hence, this thesis describes how AMoD systems are operated, what impacts can be expected of unregulated AMoD services, and how to regulate them.

1.1 State of Mobility-on-Demand Systems around 2020

In recent years, fast and mobile internet and the following digitalization enhanced the sharing economy. For large shares of the population, having access to something and using it on demand became more important than owning it. The penetration of this trend depends on the respective field; for example, business models in the media sector completely transformed, shifting away from selling hard copies (Blue-Ray, DVD, ...) of music or videos to streaming services that enable customers to enjoy their entertainment on demand. The transportation sector is also affected by this trend. However, the effect can be expected to be smaller than in media because media can be fully digitized and benefit from the advancing speed of the internet network. In contrast, the street and rail infrastructure still limit the physical transport of people and goods.

Nevertheless, the mentality of sharing is also growing in the transportation sector [S. A. S. Shaheen et al., 2017]. While owning a private vehicle used to be a status symbol, parts of the population (mainly in urban environments) can enjoy the benefits related to private vehicles without actually owning one [Zhou and D. Wang, 2019]. Mobile apps enabled new modes like car-/bike-/scooter-sharing, non-commercial ride-sharing, and ride-sourcing by transportation network companies to provide new user-centric and convenient services to millions of users. Especially transportation network companies, which offer taxi-like services, gained significant market shares and are sizable enough to create major observable impacts on cities' transportation networks. Major companies such as Didi Chuxing[1], Uber[2], Lyft[3], and Ola[4] serve millions of trips per day. This thesis focuses on such taxi-like services and uses the term mobility on demand (MoD) as an umbrella term for them. MoD services can offer

[1] https://www.didiglobal.com
[2] https://www.uber.com
[3] https://www.lyft.com
[4] https://www.olacabs.com

trips shared by multiple parties, which is denoted by ride-pooling. If an MoD service only offers exclusive trips, in which no more than one party is in a vehicle at a time, it is called a ride-hailing service.

Indisputably, MoD services provide an additional option to travelers that offer many benefits over more traditional modes of transportation from a user perspective. While taxi companies start to catch up introducing mobile phone apps, the convenience of the new MoD services with real-time information, online booking, and automatic payment quickly generated more demand than traditional taxis ever had in many cities. The results of several user surveys in different countries shows that the most occurring trip purpose is leisure (see [Tirachini, 2020] for a summary of studies and the references therein). Often, these activities take place in city centers, where parking pressure and parking prices are high. Moreover, customers can drink alcohol while enjoying the benefits of door-to-door transport [Clewlow and Mishra, 2017].

Obviously, city administrations welcome a possible correlation between MoD introduction and less drunk driving [Fortin, 2017]. However, transportation planners and city administrations have to consider multiple aspects. Depending on priorities and the attitude towards the existing modes, the perceptions of MoD systems vary and affect country/state/city regulations. Cannibalization of public transportation (PT), impact on vehicle kilometers traveled (VKT), and labor laws are among the most discussed topics.

As an example, the typical business model of Uber was declared illegal under German law, which did not address ride-hailing properly, and Uber has to stick to all existing taxi regulations (e.g., insuring trips, only hiring taxi drivers) to operate[5]. A revision of the respective law, the PBfeG, was passed in August 2021. Moreover, the actual income of drivers might be below the minimum wage [Henao and Marshall, 2019a]. However, this thesis will not go into more depth about labor laws since vehicle automation makes discussions about the business model with freelancing drivers unnecessary.

[Tirachini, 2020] also provides a compilation of several surveys analyzing MoD mode substitutions and induced trips. In these surveys, MoD users are typically asked by which mode they would have made the trip if the MoD service would not have existed. The results are somewhat different for the studied cities: for example, approximately one third would have used taxi and another third PT, while only 7 % would have used a private vehicle and 7 % of trips would not have been made in San Francisco [Rayle et al., 2016]; in Denver, only 10 % would have used a taxi, 22 % PT, approximately a third of private vehicle and more than 12 % of trips would not have been made [Henao and Marshall, 2019b]. The last-trip replacement question does not consider long-term effects on private vehicle ownership and induced travel behavior changes. Several studies in the US have studied the impact of MoD on vehicle ownership and found similar results, with approximately 10 % of users disposing of one or more cars [Rayle et al., 2016; Clewlow and Mishra, 2017; Henao and Marshall, 2019b].

Besides trip and private vehicle replacement, the impact on VKT and congestion also depends on the operating model of the MoD system. Current MoD business models employ two-sided platforms that match customers and drivers. Typically, drivers will not be able to pick up the next passenger where they dropped of the last. Since most drivers try to opti-

[5] https://www.uber.com/de/newsroom/fakten-uber-deutschland-sachlage/

mize their own revenues by driving to areas where they expect demand, the fleet control is decentralized. As a consequence, the amount of deadheading is rather large. A study based on data from Denver estimated a lower bound to be around 40 % [Henao and Marshall, 2019b]. Additionally, market competition even decreases the efficiency as deadheading generally increases with less vehicles of one MoD provider per area and driver double-apping, i.e. driving for multiple MoD providers is typically prohibited [Jiang and L. Zhang, 2018].

Pooling can achieve positive effects on VKT as multiple passengers share a ride for a part of their trajectory. However, the ratio of pooled trips in MoD services offering hailing and pooling is currently relatively small [Bansal, Sinha, et al., 2020; Moody and Zhao, 2020]. More recently, some providers, e.g., ViaVan[6] or MOIA[7], introduced MoD services only offering pooled trips to customers.

Whether these providers can operate their service profitably on their own, remains to be seen. However, as part of PT, they might not need to be profitable. Several cities started to rethink the role of MoD services and founded partnerships that typically involve subsidizing rides to fill voids in the public transit networks. Moreover, an integration of MoD and public transit apps to form mobility-as-a-service apps aim at increasing the share of intermodal trips and, in the end, private vehicle ownership.

A condition for MoD companies to receive subsidies will likely be that they can prove some social benefit. Nevertheless, evaluating these benefits is not a trivial task. For example, quantifying changes in VKT is complex as prior mode choice and induced demand also have to be considered [Tirachini, Chaniotakis, et al., 2020]. Furthermore, deriving system-wide consequences from small trip survey samples increases the uncertainty of these studies. Hence, researchers build data-driven simulation models to study their effects [Wilkes et al., 2021].

1.2 Future of Mobility-on-Demand Systems

Another field in which MoD providers could improve social welfare is sustainability and environmental impact of the transportation system, first and foremost local and global emissions. Tremendous growth of the share of private electric vehicles was achieved in China — in part by extensive regulation, for instance, by granting guaranteed licenses for electric cars in contrast to limitations for gasoline-fueled vehicles[8]. Compared to China, the market development is relatively slow in other countries. For example, the number of electric vehicles is slowly increasing in Germany even though buying electric vehicles is severely subsidized. It seems that governments have difficulties convincing European private vehicle owners to shift towards electric cars. Since operating an MoD system in cities often requires licenses from the respective administration, regulations regarding the share of electric vehicles are possible in the future. Such measures should have positive effects on the net emissions of greenhouse gases [Greenblatt and S. Shaheen, 2015]. Additionally, MoD fleet operators might want to adopt electric vehicles anyway as the economics seem to be beneficial for electric vehicles as maintenance costs are expected to be lower with an electric power train [Arbib and Seba,

[6] https://www.viavan.com
[7] https://www.moia.io
[8] https://www.bloomberg.com/news/articles/2019-02-27/in-beijing-you-have-to-win-a-license-lottery-to-buy-a-new-car

2017]. Typically, MoD trips are short-range urban trips and range anxiety will not play a role as long as the fleet operator develops an intelligent recharging strategy. The effect of electrifying the fleet on system cost and thereby on prices and customer demand for MoD is likely to remain small.

In contrast, automation of vehicles might bring considerable changes to MoD systems, both from the customer's and operator's point of view. Before elaborating on the differences brought to the MoD systems by automation, the following paragraph gives a very brief overview of the development of autonomous vehicle (AV)s. For more details and further references, interested readers are referred to the book series Road Vehicle Automation (volume 1[Meyer and Beiker, 2014] - volume 6[Meyer and Beiker, 2019]). First and foremost, it is difficult to predict when widespread adoption of autonomous vehicles will take place. In recent years, there were two major movements: the first is to build automation around existing vehicle concepts, the second is to develop new vehicle concepts for automation from scratch. On the one hand, practically all car manufacturers and their suppliers invested money into the technical extension of existing vehicle safety features (e.g., automatic cruise, lane-keeping, and lane-changing) to gradually develop self-driving capabilities. On the other hand, new players emerged on the market aiming to build entirely new vehicle concepts (e.g., Navya[9], EasyMile[10]) or provide know-how or even complete platforms for existing vehicles (e.g., Waymo[11], Mobileye[12]).

The core principles are to sense the environment with advanced technology like Lidar, combine information from various sensors and high-resolution maps to make sense of the road and surrounding objects before the autonomous vehicle plans the trajectory and realizes driving maneuvers. The complexity to achieve these steps with minimal uncertainty, which is required to ensure traffic safety, is extraordinary; for more details, the reader is referred to [Yurtsever et al., 2020] and the references therein. Nowadays, most miles are driven in quite realistic simulation environments; nevertheless, convincing the public opinion of the new technology requires real-life tests. Current test pilots in the EU employ vehicles that only drive with low speeds up to 20 km/h and need a safety driver who has to react in emergencies or after a vehicle stops because of an obstacle alongside a fixed trajectory. Highways with structural separation of oncoming traffic seem to be one of the next use cases as the complexity is limited, especially in congested low-speed situations. In contrast, mixed traffic situations with pedestrians and bicycles seem to be further away in the future.

Besides the technical development to reach a sufficient level of safety, clarifying liability issues represents another critical issue to reach higher automation levels [Maurer et al., 2016]. Legal frameworks differ all over the world and – without drivers – the manufacturers will likely be liable in some form. Hence, widespread adoption of AMoD might still be further in the future. However, Waymo launched an AMoD service called Waymo One in 2019, which operates in Phoenix, Arizona. Besides vehicles with safety drivers, it also operates fully autonomous vehicles within a geofenced area, thereby reaching level 4 automation[13]. Even in Germany, several companies announced a near-future implementation of autonomous vehicles

[9] https://navya.tech/en/
[10] https://easymile.com/
[11] https://waymo.com/
[12] https://www.mobileye.com/
[13] SAE definition, see https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety

in mobility services. SIXT started a cooperation with MobilEye to start in 2022[14] and MOIA plans to use self-driving VW vehicles in 2025[15].

After this discourse on the current state of vehicle automation, the impacts of automation on MoD systems are discussed. From a customer's perspective, the system will be rather similar: a customer requests a trip per app by providing the origin and destination of the desired trip. Possibly, the request also contains constraints on the pick-up and drop-off time. The operator provides information about the service AV to the customer before it arrives at the pick-up location. Then the customer enters the vehicle and becomes a passenger until she disembarks the AV at the destination. While not having a human driver might be the largest emotional change, the pick-up will be most different process-wise. Since there is no human driver to confirm the identity, confirmation systems like those developed for today's carsharing systems are imaginable. In general, AMoD could also be viewed as a development of free-floating carsharing systems under automation. The difference to today's services would be that customers no longer walk to the car, but the AV drives to them. Therefore, the systems evolving from ride-hailing and carsharing become hardly distinguishable at first glance.

Even though the evolution of classical MoD systems to AMoD remove most differences between them, various system designs are still possible. For example, a system evolving from a carsharing business could still allow users to choose an AV from a set of idle fleet vehicles, which might contain different vehicle models. In such carsharing system with AVs, customers might even want to drive from their origin to their destination manually while enjoying the comfort of an autonomous vehicle pick-up trip. In this case, the customers would not have to communicate their destination to the operator (as in today's carsharing systems). Similarly, AMoD based on today's ride-hailing systems could still use the two-sided platform approach with AVs. However, it seems tedious for private AV owners to still accept every single task manually instead of just allowing the AMoD operator control of the AV for a certain amount of time. an operator would have to consider legal, technical, and safety issues when using private AVs. Therefore, it will probably be easier for an AMoD business to have complete control over the AVs (whether they buy or lease the AVs).

In general, central control is advantageous over a system with individual agents, i.e., customers or vehicle drivers/owners who perform (local) optimal decisions for their benefit. Contrarily, central control allows global optimization with respect to a system objective. [Dandl and Bogenberger, 2018] studied the difference between the system with decentralized control, in which customers choose their vehicles, and a centrally controlled fleet. Their case study in Munich, Germany, showed that central control decreases the number of not served customers by more than half. Additionally, central control can reduce the share of empty VKT by approximately 45 % or 20 % for fleet sizes of 3000 or 4000 vehicles, respectively.

In all likelihood, AMoD services will be much cheaper than today's MoD systems [S. Shaheen, 2018]. The most important cost component of ride-hailing and taxi services, the driver, will be replaced by additional investment costs for the higher technological standard. However, as the technology develops, these investment costs are expected to decrease, making AMoD drastically cheaper than current MoD services. Furthermore, insurance, maintenance, and

[14] https://www.sixt.de/magazine/news/robotaxi-mobileye-sixt/

[15] https://www.moia.io/en/news-center/vwcv-moia-and-argo-ai-present-roadmap-for-autonomous-ride-pooling-in-hamburg

energy consumption costs will probably decrease as a consequence of electric AVs [Arbib and Seba, 2017; Bösch et al., 2017]. The prices of today's carsharing services can serve as an upper bound since they do not include driver costs. Nevertheless, AMoD is likely to be much cheaper because the utilization of vehicles can be much higher in AMoD systems [Spieser et al., 2016; Dandl and Bogenberger, 2019]. The effective area in which a vehicle is available is much larger as it is determined by walking speed and acceptable walking time in a carsharing system and driving speed and acceptable waiting time in an AMoD system. Then again, vehicles are likely to require more cleaning and maintenance processes than current carsharing vehicles. Depending on the frequency and costs of cleaning, these costs could account for \$0.05 to \$0.1 per vehicle-mile [Litman, 2019], which is considerable, but small compared to driver costs. Additional investment costs need to be considered if AMoD operators require their own charging infrastructure [T. D. Chen and Kockelman, 2016]. [Burns et al., 2013] initially estimated AMoD costs around \$0.15 - \$0.50 per mile. Assuming similar AMoD business and vehicle costs than today's carsharing service, [Dandl and Bogenberger, 2019] estimated costs of 0.25 € per km (approximately \$0.44 per mile). Approaches considering the single cost components estimated per-mile costs of \$0.15 [Arbib and Seba, 2017], \$0.45 [T. D. Chen and Kockelman, 2016], \$0.45 [J. Walker and Johnson, 2016], \$0.68 (converted from 0.43 CHF/km) [Bösch et al., 2017], and \$0.80 [Litman, 2019]. Even though these estimations differ quite a lot, all have in common that the prices will be much lower than today's prices for ride-hailing and taxi services, which are in the range of \$1.50 to \$2.50 per mile [Litman, 2019]. Hence, AMoD systems could achieve a considerable increase in market share and research about these systems is highly relevant. Due to these lower prices, the popularity and demand of AMoD systems can be expected to eclipse today's systems, thereby potentially disrupting the transportation system.

The introduction of AVs potentially brings another far-reaching phenomenon: privately owned AVs. Households with such AVs could utilize the autonomous driving capabilities to send the vehicle from one household member to the other. Optimizing the AV route for the planned activities allows the household to replace multiple conventional private cars but can increase VKT significantly [Correia, Gonçalo Homem de Almeida and van Arem, 2016; X. Xu et al., 2019]. Depending on household size and activities, additional empty VKT of 34 to 51 % were found in a study by Xu et al. [X. Xu et al., 2019]. Centrally coordinated AMoD fleets also require empty VKT to travel between customers, but the ratio of distances driving without and with passengers will be better. For example, a private AV would have to return home after a trip to work to serve another family member, resulting in 50 % empty VKT. For the same trip, this family member could use a nearby AMoD vehicle, which likely will be much closer, thereby producing less VKT without passengers. Therefore, [Narayanan et al., 2020b] conclude that AVs should be introduced in shared systems from a policy perspective.

Because of the possibly disruptive changes caused by vehicle automation, transportation planners and regulating authorities should act ahead of time. Since introducing regulations typically is a long-lasting process, they should forecast the consequences of an unregulated introduction of AVs, develop possible regulations, and analyze their impacts. It cannot be expected that there will be one global solution that works for every use case. The financial viability is quite different in rural and urban areas. Furthermore, the effect of empty trips

might hardly affect the traffic quality in rural areas whereas the vast amount of extra VKT produced by private AVs is likely not acceptable in cities that already have congestion on their street networks. Moreover, the impact on the PT system or even possible extensions of the PT with AMoD will depend on the available PT system itself: well-structured PT systems may be extended by first/last mile services or new services might be offered in rural areas, where no PT option was viable before, but AMoD can also be expected to attract demand from existing PT systems. This thesis focuses on urban areas since negative externalities and challenges are expected to be larger. Furthermore, this thesis assumes that private AVs will not be viable for a significant amount of the population (either they are too expensive or prohibited) and will not affect urban transportation networks.

Developing and analyzing regulations for AMoD systems involves a new feature for transport planners. Typically, there is no intermediary between the planned measure and the traffic demand for applications like road planning, intersection control or even public transportation planning. Often, measures are directly planned to achieve a certain level of service for a given demand (one-level optimization). Transport planners also utilize frameworks with a feedback loop (bi-level optimization) if the measure is expected to change demand. However, fleet operators, especially private AMoD providers, add a new layer to the problem. City administrations try to introduce regulations that optimize a social welfare measure (that they define based on their priorities). As a new element, AMoD providers will adapt their service design and fleet operation to maximize profit. Finally, travelers will choose among their mobility options, thereby determining the effects on the whole transportation network. These traveler choices depend on the regulations, the offered AMoD service, and the other mobility options of the respective travelers.

1.3 Research Objectives and Methodology

The overarching goals of this thesis are to create a better understanding of AMoD operations and to provide an optimization framework for regulatory measures in the presence of AMoD systems. Insights into operational aspects of AMoD systems are critical to estimate the impact of regulations on the AMoD service design and operation and thereby on the service offered to customers and non-customers of the AMoD service. Today's pilots do not provide the necessary scale concerning fleet size, vehicle speed, and fares. Hence, simulation models are essential as analyzing tools because (i) fully scaled AMoD systems are not yet available and (ii) simulation is an excellent tool to analyze "what-if" scenarios. The latter becomes even more crucial when studying regulations considering long-lasting administrative processes of real-world implementations.

This thesis aims to answering research questions on the operation of AMoD systems and the regulation of transportation systems with AMoD services. The main operational questions in this thesis are as follows:

RQ 1.1 How can an AMoD fleet be operated effectively and efficiently?

RQ 1.2 Which are the most important operational variables?

To study possible regulations, the AMoD service has to be viewed in the context of the overarching transportation system. Following questions are guiding this thesis:

RQ 2.1 What are the effects of an AMoD service on a transportation system?

RQ 2.2 How will an operator react to different regulatory measures?

RQ 2.3 How can city administrations regulate the transportation system considering the reaction of AMoD operators?

The subsequent chapter surveys existing literature on AMoD systems. After that, chapter 3 and chapter 4 answer the operational and regulatory research questions, respectively. The respective chapters describe mathematical problem formulations, operational and regulatory strategies and indicators measuring the performance. Figure 1.1 illustrates the approach taken in this thesis.

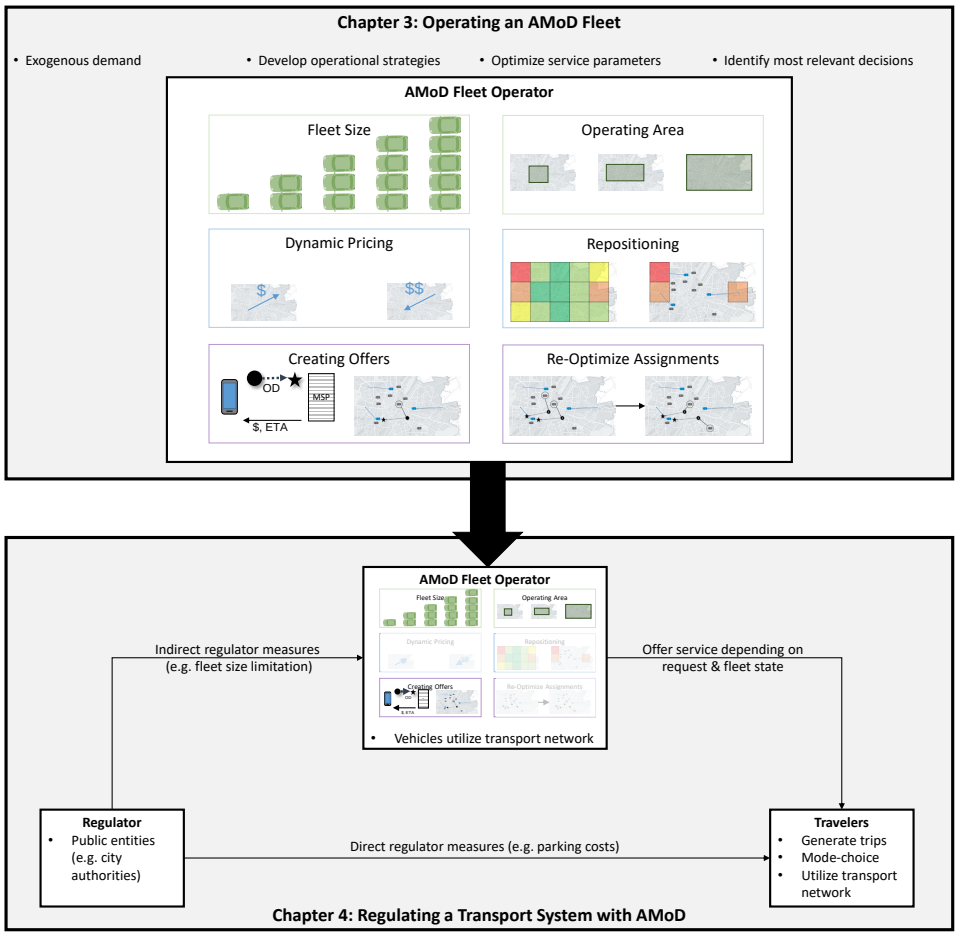


Figure 1.1: Overview of Research Methodology

In the chapter focusing on the operation of an AMoD fleet, exogenous demand is assumed to compare the impact of several strategies and decisions. These decisions can be viewed on different time scales; as an example, choosing a fleet size and operating can be regarded as long-term planning processes, whereas dynamic pricing, repositioning, creating offers and re-optimizing assignments have to be performed much more frequently.

The insights gained from this chapter as well as (parts of) the model to operate an AMoD fleet are integrated into a larger context in chapter 4. The interactions and effects of decisions made by a regulating entity, the AMoD fleet operator, as well as travelers are studied.

As modeling detail and computation time are limited, the framework to study AMoD regulations in chapter 4 can accommodate only a rather small number of operator variables. Since the parameter space of decision variables grows with more advanced operator strategies, this thesis contains a separate case study on the operation of AMoD systems to identify the most relevant operator variables at the end of chapter 3 and keep these as degrees of freedom for the AMoD operator in chapter 4. Finally, chapter 5 will summarize and discuss the main contributions of this thesis, integrate them into the super-ordinate context, and outline future work.

Chapter 2

Literature Review

AVs and AMoD are still in the pilot phase, having a relatively small impact on the transportation systems. Therefore, the expected consequences need to be explored in theory, i.e., via stated preference experiments, analytic approaches, and simulation. Many recent studies have applied these methods to generate insights about scaled AMoD systems. As this thesis focuses on the operational aspects, the subsequent sections will focus on the literature about the supply side related to AMoD fleets. For more information about the demand side, land use, traffic control, traffic safety, or economic stakeholders (e.g., car manufacturers, suppliers, tech and infotainment companies), the reader is referred to Duarte and Ratti [2018], Gkartzonikas and Gkritza [2019], and Narayanan et al. [2020b] and references therein for further information.

AMoD services can vary in several components. The two main components are service design and fleet control. Service design defines the system as it presents itself to its customers, while fleet control determines how an operator can achieve this design. After touching on AMoD service design, this chapter reviews various aspects of modeling AMoD systems to address operational policies/strategies, studied in different depth in the literature that vary from macroscopic models to real-time implementations. The analyzed sub-categories of *AMoD Service Design*, *AMoD System Modeling*, and *AMoD Fleet Operation* are shown in Table 2.1.

Service Design	**System Modeling**	**Fleet Operation**
- operating area	- network model	- fleet sizing
- charging infrastructure	- customer & vehicle model	- charging strategy
- hailing/pooling	- travel times & - routing	- customer assignment strategy
- fleet composition	- AMoD competition	- repositioning strategy
- pricing considered	- AMoD demand	- dynamic pricing strategy
- customer interactions	- stochasticity of interaction	
- time constraints	- PT cooperation model	
- MaaS & - PT integration		

Table 2.1: Overview of highlighted topics in literature review.

A similar taxonomy (albeit split into typical vehicle routing problems (VRP) and new AMoD categories) for the operation of AMoD systems was created by M. Hyland and Mahmassani [2017]. Furthermore, Jing et al. [2020] created another recent literature review article about agent-based simulations of AVs.

2.1 AMoD Service Design

Service design determines the kind of service an AMoD provider wants to offer and advertise. AMoD providers are likely to keep this design for long periods of time to build a brand and provide a certain customer experience; hence, planning the service design is a crucial step for AMoD providers. The service design models define the operating area of the AMoD service and the available charging infrastructure. Moreover, they determine whether a hailing or a pooling service is offered, whether the fleet consists of one or multiple vehicle types, how the service should be priced, how the customer interaction is defined, whether there are time constraints on pick-up and drop-off, and whether there exists some integration into the PT system. The subsequent paragraphs provide an overview of how these topics were addressed in literature. Table A3 in the Appendix D classifies papers that put part of their focus on service design or whose service design model differs in at least one category from most other publications.

Operating Area: Service is likely to be offered in certain operating areas, which means that both the pick-up and drop-off points have to be within a particular geofence. Due to higher vehicle utilization, these can be expected to be larger than today's carsharing geofences [Dandl and Bogenberger, 2019]. Smaller areas require fewer vehicles, and vehicles will have shorter trips, allowing one vehicle to serve more trips [Fagnant and Kockelman, 2014]. Service providers will likely base geofences on network topology (as around the outer highway belt in [Dandl, Bracher, et al., 2017]), geographic attributes (such as a certain length and width [Fagnant, Kockelman, and Bansal, 2015]), or administrative boundaries (such as Manhattan borough [Marco Pavone, 2015; Dandl, M. Hyland, et al., 2019] or the central business district in Singapore [Azevedo et al., 2016]). Nowadays, MoD providers often use socio-demographic and point-of-interest data to estimate demand for their systems and to determine whether an area should be part of the operating area. Engelhardt, Dandl, Bilali, et al. [2019] analyzed expected trip data starting in a given set of zones within a large operating area; in an iterative procedure, they removed zones with the lowest number of trips and updated the trip data by removing trips ending in eliminated zones. In literature, there are also simulation-based optimization approaches to define the service area [Liang, G. H. d. A. Correia, et al., 2016; Bischoff, Kaddoura, et al., 2018]. An evaluation of the relative trip number and remaining operating area determined the final operating area. As in carsharing, there could be a station-based operation of AMoD [Spieser et al., 2016; R. Zhang and M. Pavone, 2016], but removing the option of the vehicle coming (closer) to customers is likely to make the service less attractive (see a comparison of TNC and carsharing market shares [Clewlow and Mishra, 2017]). A solution between free-floating and station-based is a network of virtual stop points, which serve as meeting points for customers that are pooled together. This aggregation of customers helps to enhance the performance of carpooling [Stiglic et al., 2015] and ride-pooling services [Zwick and Axhausen, 2020; Engelhardt and Bogenberger, 2021; Andres Fielbaum et al., 2021]. Among other, different qualities of street and public transport networks, demand patterns, demand densities are a consequence of the choice of different operating areas. These all impact the AMoD system and also its effect on the transportation system. On the one hand,

this necessitates case studies with different operating areas, and on the other hand, it makes a comparison of different operating strategies rather difficult. Most studies analyze AMoD systems in urban environments, but there are some studies that include more rural regions, for instance, [Boesch et al., 2016; Wen, Y. X. Chen, et al., 2018; Vosooghi, Puchinger, Jankovic, et al., 2019; Sieber et al., 2020; Wilkes et al., 2021; Zachariah et al., 2014]. Some operations research focused papers study their algorithms on existing benchmark instances [Bongiovanni et al., 2019; H. Zhang et al., 2019] or new test instances [T. D. Chen and Kockelman, 2016; Wenwen Zhang et al., 2015; J. J. Q. Yu et al., 2018] to address the issue of comparability. Sheppard et al. [2019] collected data from 9 cities and the US to apply their approach to a larger number of cases. Oke, Aboutaleb, et al. [2019] started another project to produce more general results and insights by classifying city types and then studying the classes by case studies of the archetype cities [Oke, Akkinepally, et al., 2020].

Charging Infrastructure: In order to provide an AMoD service with electric vehicles, charging infrastructure will be required. Some studies used publicly available charging station data [Dandl and Bogenberger, 2019; Fehn et al., 2019; L. Li, Pantelidis, et al., 2021], including both the position of charging stations and the number and power of charging units. The power of charging units varies between 1.5 kW [Liang, G. H. d. A. Correia, et al., 2016] and 300 kW [Boewing et al., 2020], and several studies test the sensitivity of this parameter on the AMoD system. Models to determine the charging infrastructure are necessary when no data is available or when the available infrastructure is insufficient to serve hypothetical AMoD services. T. D. Chen and Kockelman [2016] developed a methodology to generate charging infrastructure during warm-up simulations. Jung, J. Y. Chow, et al. [2014] introduced a two-level optimization framework, where the amount of charging units per station was derived in the upper level while the lower level contained a simulation of the fleet operation. Similarly, Kang et al. [2015] developed a multi-level framework, where the choice of charging station locations in one layer served as input for the fleet assignment layer. A heuristic to reduce the number of charging locations was employed by Bauer et al. [2018] between simulations to create their charging infrastructure distribution. H. Zhang et al. [2019] derived a methodology for systems spreading over multiple cities where charging station locations result from a mixed-integer linear problem. Another optimization approach based on the maximal-covering location problem was utilized by Vosooghi, Puchinger, Bischoff, et al. [2020]; additionally, the authors considered battery swapping in their study.

Hailing/Pooling: A service provider has to determine whether they want (i) customers to always have a vehicle for themselves (hailing), (ii) customers to accept pooling automatically, or (iii) customers to have the choice whether they want hailing or accept pooling. While hailing is the predominant service type in today's MoD services, pooling of customers will be needed to reduce the load on the street networks and be operated more efficiently [M. Hyland and Mahmassani, 2018b]. Simulations of AMoD systems in literature typically assume operation in options (i) or (ii). Whether or not an AMoD provider offers pooling influences the vehicle choice as pooling services need vehicles that can accommodate multiple passengers. Therefore, pure pooling services typically employ minibuses or vans. Studies

assume different vehicle passenger capacity limits, e.g., two [M. F. Hyland and Mahmassani, 2020], four [Jung, Jayakrishnan, and Park, 2016; Engelhardt, Dandl, Bilali, et al., 2019], six [Zwick and Axhausen, 2020], or even up to ten [Alonso-Mora, Samaranayake, et al., 2017]. In simulations, it is possible to remove the vehicle capacity constraint and evaluate the maximum vehicle occupancy [Fagnant and Kockelman, 2018; Ruch, Lu, et al., 2020] to find a good composition of vehicle sizes.

Fleet Composition: Vehicle model type, which implicitly include number of seats, battery size, and energy consumption, should match the operational concept. Most current MoD systems have a mixed fleet because freelance drivers utilize their own private vehicles. On the contrary, the ride-pooling service MOIA prefers a homogeneous fleet with its own branding. Most studies do not differentiate between vehicle models or explicitly assume homogeneous fleets. Niels and Bogenberger [2017]observed different perceptions of an electric vehicle and a vehicle with internal combustion engine in a study analyzing carsharing app and rental data. Dandl and Bogenberger [2018] studied a fleet with two vehicle models and found that customers preferring one over the other increases VKT compared to scenarios with a homogeneous fleet of the same size. For two vehicle types, pick-up trips to a specific vehicle type are longer on average as the density of available vehicles of specific models is smaller than the density of available vehicles in the whole fleet. With respect to the electrical properties of vehicles, the sensitivity of the AMoD system to different battery sizes is analyzed by [T. D. Chen and Kockelman, 2016; Bauer et al., 2018; Sheppard et al., 2019; L. Li, Pantelidis, et al., 2021; Vosooghi, Puchinger, Bischoff, et al., 2020]. More complexity is required to analyze mixed fleets. Zhu et al. [2017] assumed an infinite pool of right-sized vehicles at each stop,Wallar, Schwarting, et al. [2019] utilized a similar concept for an offline tool to determine the required fleet composition, Y. Liu, Bansal, et al. [2019] applied a Bayesian optimization framework to find the correct fleet from iterative simulations, and Atasoy et al. [2015] developed an operational model, where a set of homogeneous vehicles dynamically changes its effective passenger capacity based on the use-case.

Pricing: The price system directly impacts the profit an AMoD operator makes. The price level is a key determinant for the attracted demand. The AMoD service could be integrated in the public transportation fare system [Shen et al., 2018] or set its own prices. Similar to today's MoD services, fares for AMoD services can have multiple components. Typically, there is a customer trip length and/or duration component. Moreover, AMoD providers might want to introduce a base fare or additional distance fare for the pick-up trip to compensate for the empty trips made by fleet vehicles and attract less short trips that were previously walked or driven by bike [Wilkes et al., 2021]. Additionally, fare reduction and fare splitting mechanisms are possible in ride-pooling: customers could receive a discount for their willingness to pool or be compensated for the amount of detour they had to endure [Kucharski and Cats, 2020; Wen, Y. X. Chen, et al., 2018]. Since customers typically want to be informed about trip prices ahead of their trip, the former is easily implemented while the latter would likely require an upper bound for the fare representing the direct route in the information stage and the final payment after the trip and the detour have been realized. In theory, realistic pricing

will be the result of an optimization process [Sebastian Hörl et al., 2019]. However, as large-scale AMoD systems are not operated yet, there is value in studies trying to estimate the fixed and operational costs as output to provide a first indication [Bösch et al., 2017; Dandl and Bogenberger, 2019]. These estimates also included variable electricity costs in some cases [Fehn et al., 2019; Turan et al., 2020]. Furthermore, dynamic pricing can be considered by AMoD providers. Dynamic pricing can help to maximize current and future profit as in other fleet operations [Figliozzi et al., 2007; Lippoldt et al., 2018; Lippoldt et al., 2019]. In the temporal dimension, higher prices can make more profit out of the remaining capacity in times of high vehicle utilization [Al-Kanj et al., 2020; Turan et al., 2020]. Spatially varying prices can help to increase the utilization of vehicles that would remain without a customer in low-demand zones. Some of these additional trips might even help to balance the fleet and reduce empty repositioning [Al-Kanj et al., 2020] or limit the flow on certain street sections [Salazar, Rossi, et al., 2018]. With multiple operators, fares could also be dynamically generated based on auctions [Kleiner et al., 2011; J. J. Q. Yu et al., 2018].

Customer-Operator Interaction: It is likely that — as in MoD systems nowadays — applications using mobile internet will be the means of customer-operator interaction in AMoD systems. The design of this interaction shapes the VRP of the fleet operator. In literature, the following models were found: (i) a customer request can be binding, i.e., is a booking, and has to be served by the operator at some point [M. Hyland and Mahmassani, 2018a], (ii) a customer requests a ride but leaves the system if she is not picked-up before a certain time [Spieser et al., 2016], (iii) a customer requests a ride but operators can accept/reject a customer right away [Dandl, M. Hyland, et al., 2019], and (iv) a customer requests service offers that the customer can accept or reject [Dandl, Bogenberger, and Mahmassani, 2019; Al-Kanj et al., 2020; Wilkes et al., 2021]. In the last case, the operator also has the option to make no offer. Cases (ii) and (iii) are closely related as an operator can estimate whether a vehicle can serve a customer within a specified waiting time. The equivalency of both cases can be proven under certain assumptions [**Dandl.2021a**]. Additionally, it is possible that multiple service offers are connected in a mobility-as-a-service app [Atasoy et al., 2015; J. J. Q. Yu et al., 2018] or customers retry to be served at a later point in time [Dandl, Grueber, et al., 2019]. Moreover, the exchanged information flow can vary for different system designs. Less specific information, like a pick-up time window, can be communicated at the time of the offer (in an immediate response system) and specified later on (when an assignment is locked) [Dandl, Grueber, et al., 2019; Erdmann, Dandl, and Bogenberger, 2021]. Other service design aspects, which are typically not modeled in simulations because of expected minor global impacts, are asynchronous information flow or delays in information [Bertsimas et al., 2019; Dandl, Bogenberger, and Mahmassani, 2019; J. Yu and M. F. Hyland, 2020], cancellation policy and penalties, and maximal vehicle idle duration in case customers show up too late or not at all for pick-up [Andrés Fielbaum and Alonso-Mora, 2020].

Time Constraints: Customers or service providers can set constraints on pick-up and drop-off times. First of all, service providers could limit operating hours but are unlikely to do so since

AVs are available 24/7 and do not have to follow labor regulations. Customer time constraints are connected with the customer-operator interaction. In a system with a service guarantee, i.e., an AMoD system where the operator has to serve each booking, hard constraints would enforce a very large fleet size. On the other hand, if there were no time constraints, a few customers in regions with low vehicle availability would wait a very long time. This effect would probably cause low customer acceptance. To avoid these unacceptably long waiting times, soft constraints, i.e., penalty terms in the objective function, are utilized, e.g., by M. Hyland and Mahmassani [2018a] and M. Hyland, Dandl, et al. [2019]. Explicit drop-off time constraints are typically only implemented for pooling services since vehicles will take the fastest path between pick-up and drop-off location in hailing services, thereby implicitly constraining the drop-off time. Detour constraints can be implemented in various ways, e.g., latest arrival time [Dandl, Grueber, et al., 2019], maximal absolute detour time [Santi et al., 2014], relative detour time [Engelhardt, Dandl, Bilali, et al., 2019], a combination of these [Zwick and Axhausen, 2020], the derivation from an initial schedule [Atasoy et al., 2015], or a certain number of common destinations [Zachariah et al., 2014]. In an on-demand service, the customer's pick-up is usually constrained by a certain maximal waiting time, but time windows can also be applied [Engelhardt, Dandl, Bilali, et al., 2019; Erdmann, Dandl, Kaltenhaeuser, et al., 2020]. A maximal wait time constraint is actually a special case of a pick-up time window in which customers want to be picked up immediately, and therefore, the minimum waiting or reservation time is 0. Contrarily, if customers do not want to be picked up right away, the system is denoted reservation-based [Bilali, Dandl, et al., 2019a]. There can also be constraints on the combination of waiting and in-vehicle travel time, denoted by total travel delay [Martinez et al., 2015] or the time customers wait for information [Dandl, Grueber, et al., 2019; S. Wang et al., 2019] before they decide to drop their request. While time constraints in literature are commonly treated homogeneously among all users, they can also be heterogeneous, i.e., specific for each AMoD request. For instance, Kucharski and Cats [2020] derive the the time constraints in their model from travel utility functions.

MaaS/Intermodal PT Integration: The interaction of customers and operators could also be indirect. Travelers could use so-called mobility as a service (MaaS) apps that bundle multiple mobility options in one application [Atasoy et al., 2015; Kamargianni and Matyas, 2017; Jittrapirom et al., 2017]: after communicating origin, destination, and planned departure/arrival time, the MaaS app queries routing services (to compute the private vehicle trip duration), the public transportation provider (for the best public transportation route), and MoD and AMoD services (for their respective offers). Additionally, MaaS applications might be able to coordinate and combine different modes to create seamless intermodal trips [S. Shaheen and Chan, 2016]. Several studies focused on the AMoD application as a first/last mile feeder system, e.g., [Vakayil et al., 2017; Shen et al., 2018; Wen, Y. X. Chen, et al., 2018], but network-wide integration is possible [Salazar, Rossi, et al., 2018]. The introduction of AMoD services are most effective for the transportation system if the AMoD and PT systems are designed together [Pinto et al., 2019] or the AMoD service is explicitly offered in areas with bad PT connection [Dandl, Grueber, et al., 2019].

2.2 Modeling of AMoD Services in Transportation Systems

The service design defines the AMoD service from the customer viewpoint. However, the performance of that service is controlled by the operational strategies that an AMoD provider applies. To evaluate these strategies from an operational perspective or evaluate the impact of an AMoD service from a transportation system perspective, models are required. These models cannot entail all details that are relevant for a real-world AMoD service, but should reflect the most important aspects. Both human and computational resources in the data collection and model building processes limit the level of detail. On the one hand, more detail is required to analyze specifics about AMoD services; on the other hand, tractability is easier for less detailed models. Hence, there is a trade-off that researchers have to consider when they design their models. The subsequent paragraphs summarize modeling approaches to the most relevant topics for AMoD systems and Table A4 (in Appendix D) classify a selected list of papers. The same classification is available for an extended list of papers in the appendix. The column *study focus* in these tables is added (to the topics described below) to allow a quick impression of whether the study addresses only operational issues (O), includes demand modeling (D), or considers traffic modeling (T) in more detail.

Network: Transportation systems are typically modeled as networks: stations and their connections naturally build the nodes and edges of the PT network; the street network can be created by defining sections as edges and intersections as nodes. Using zone centroids as nodes and their connections as edges allows the definition of more aggregated networks. The choice of spatial resolution between nodes and thereby the total number of nodes can be critical for computational effort. Some papers, e.g., [M. Pavone et al., 2012; Spieser et al., 2016; Carron et al., 2019], reduce the network size to a rather small number of stations, representing real stations (like carsharing stations or taxi stands) or all demand inside a zone and connections between them. This approach is valid for station-based AMoD systems. Moreover, it can be utilized for studies focusing on other aspects than the actual user-vehicle assignment process or as an abstract network layer for fleet control processes [Azevedo et al., 2016; Ruch, Horl, et al., 2018; Q. Li and Liao, 2020]. Fagnant and Kockelman [2014], Fluri et al. [2019], and Al-Kanj et al. [2020] even defined a hierarchical cascade structure for different spatial aggregation levels. The benefit of cheap routing computations also motivated grid-shaped networks [Fagnant and Kockelman, 2014; T. D. Chen and Kockelman, 2016], in which routing can be reduced to finding shortest distances according to the Manhattan metric. The idea of using Euclidean [Burns et al., 2013] or Manhattan metric [M. Hyland and Mahmassani, 2018a; Dandl, M. Hyland, et al., 2019] for routing also allows the use of continuous planes, which allows a more realistic spread of demand than a grid network, but excludes modeling network effects like congestion. Thanks to the availability of open data source projects like OpenStreetMap (OSM)[1], the amount of work (for a single researcher) to build networks was reduced significantly in the past 15 years. Therefore, many studies can utilize realistic street networks, on which demand is defined and

[1] https://www.openstreetmap.org

the AMoD vehicles can move.

Customer & Vehicle Model: Agent-based models allow a realistic representation of vehicle and customer movements and boarding processes. Therefore, operational models have to include an assignment process of customers to vehicles; any optimization process of these assignments is bound to contain integer variables. These models are computationally rather expensive as the actions of many agents have to be considered. On the plus side, agent-based modeling is often more straightforward as processes are modeled rather than possible effects in macroscopic models. Some studies employ flow models in which customers and vehicles traveling on the same network edge are aggregated. For better computational performance, these studies assume continuous flows, i.e., the integrality constraints are dropped in optimization problems [M. Pavone et al., 2012; Levin, 2017]. For real-life applicability, a quantization of continuous flows is required for such approaches [R. Zhang and M. Pavone, 2016]. The results from both agent-based and network flow models are bound to be specific to the modeled operating area. Macroscopic models remove even more detail thereby trading off exactness in the analysis for a certain use case to gain more general insights. Several macroscopic models were developed for ride-pooling, e.g., [Tachet et al., 2017; Daganzo and Ouyang, 2019; Herminghaus, 2019; Bilali, Engelhardt, et al., 2020]. From a modeling perspective, creating useful macroscopic representations is quite challenging and often requires symmetry assumptions like homogeneity in space [Bilali, Dandl, et al., 2019b; Bilali, Dandl, et al., 2019a]. Macroscopic MoD models are also useful to represent complicated interactions in two-sided platforms of today's MoD systems [Z. Xu et al., 2019; Nourinejad and Ramezani, 2019] or estimate future states in model-predictive control approaches [Hyytiä et al., 2012; Dandl, Fehn, et al., 2020].

Travel Times & Routing: The AMoD operator will decide where to send its vehicles based on the current state of the transportation system. Furthermore, vehicles move according to the assigned routes in between routing decisions. Most studies utilize deterministic frameworks, in which the expected travel time (used to make routing decisions) and the realized travel time (the amount of time the vehicle actually needs for the route) are the same. Exceptions are models with integrated traffic simulation [Marczuk et al., 2015; Dandl, Bracher, et al., 2017; Levin et al., 2017; Levin, 2017], which inherently generate stochastic travel times. Deterministic but time-varying travel times are more common. Typically, the travel times are generated from a traffic simulation, e.g. [Dandl, Grueber, et al., 2019], or based on OSM free-flow velocity combined with real traffic data [Vazifeh et al., 2018; Dandl, M. Hyland, et al., 2020; Zwick and Axhausen, 2020]. Some studies also ignore the congestion aspect and use free-flow travel times throughout the day. Models without an explicit network connect two points in a plane by the shortest path in its metric (Manhattan/Euclidean distance) [Daganzo and Ouyang, 2019; Dandl, Fehn, et al., 2020]; constant speed throughout the operating area is assumed if travel times are necessary [M. Hyland and Mahmassani, 2018a; Herminghaus, 2019].

With respect to routing, grid networks are similar to a plane with Manhattan metric; hence constant speeds are assumed for these networks in most cases [Wenwen Zhang et al., 2015; T. D. Chen and Kockelman, 2016; Zhu et al., 2017]. Most approaches in AMoD

literature assume that vehicles will use the fastest path between two stops. The routing problem is to find the path, i.e., the sequence of edges in a network, that minimizes a certain cost function, typically the shortest or fastest route. This problem is typically formulated as an integer linear program. Cost functions with path-level components generally require an enumeration of all feasible paths and are computationally inefficient. For cost functions that are the sum of edge-level costs, dynamic programming [Bertsekas, 2005] enables routing algorithms that are computationally very efficient. The oldest and most famous one is Dijkstra's algorithm [Dijkstra, 1959], which builds a frontier of nodes to search and always adds the neighbors of the current lowest-cost node of the frontier until the destination node is reached. This approach leads to radial shapes of the explored nodes around the origin node. In order to reduce the number of explored nodes and thereby increase the computational efficiency, several algorithms, e.g., bidirectional search, A^* [Hart et al., 1968], and ALT [Gutman, 2004], have been introduced. These algorithms adapt the cost function to aim the search in the direction of the destination. With the availability of better computers (CPU and memory), more advanced routing algorithms were developed that preprocess networks[2] to achieve high-speed online routing queries. The reader is referred to Delling et al. [2017] for a recent survey. Of course, saving the complete travel time matrix in memory, which is possible for some urban networks, achieves the best computational performance.

Typically, routing a fleet of vehicles is broken down to giving tasks/destinations for each vehicle and each vehicle using these routing algorithms to find its best route. Nevertheless, some studies investigate the potential of coordinated fleet routing, where AMoD vehicles are distributed among different routes to minimize congestion in the network. Even though conceptually interesting, significant system-wide performance benefits require a large market penetration and agent-based approaches seem to be lacking scaling properties [Liang, G. H. d. A. Correia, et al., 2018]. The utilization of time-dependent marginal travel times from system-optimal traffic assignment [Peeta and Mahmassani, 1995] might make them more efficient. Flow models are more efficient to address congestion-aware routing [Rossi et al., 2018; Salazar, Tsao, et al., 2019] but require a quantization to be applicable for real-world applications.

PT Cooperation Model: When an AMoD and line-based PT system are combined, the number of intermodal route combinations between an origin and a destination pair becomes easily intractable. Salazar, Rossi, et al. [2018] and Zgraggen et al. [2019] connect the PT station with some waiting/walking edge to the street network and allow all possible routes, which is computationally feasible due to their flow-based vehicle and customer model. To work with actual customer- and vehicle-route assignments, Vakayil et al. [2017] and Pinto et al. [2019] preprocess the fastest transit-AMoD hyperpaths for each origin-destination pair. Another approach to limiting the solution space is delineating the catchment area where an AMoD feeder serves PT stations [Basu et al., 2018; Ma et al., 2019; Oke, Akkinepally, et al., 2020]. The seamless transfer between both transportation modes requires more research. On the one hand, a guarantee for on-time arrival by the AMoD service in the first-mile problem would need attention as current studies do not consider these constraints explicitly. On the

[2] i.e., perform a lot of routing computations offline, e.g., to build contraction hierarchies [Geisberger et al., 2008]

other hand, a reservation for the last-mile problem could already be made at the beginning of the PT trip leg instead of creating the request at the arrival time at the final PT station [Ma et al., 2019].

AMoD Competition: Most studies assume a monopoly of one AMoD provider. Some exceptions consider the competition between AMoD providers. Dandl, Bogenberger, and Mahmassani [2019] simulated two competing operators with different fare systems and showed the impact on the AMoD performance metrics. Simonetto et al. [2019] introduced a concept in which the AMoD operators provide the marginal change of the cost function related to a customer insertion to a central agency, which considers these marginal changes and a predefined market share for the companies in its decision to keep the fleet sizes stable and vehicles utilized. Atasoy et al. [2015] and Y. Liu, Bansal, et al. [2019] analyze AMoD systems offering hailing and pooling simultaneously that are essentially in competition. From an efficiency viewpoint, the absence of competition can be beneficial as AMoD services have positive scaling properties, which are countered by market fragmentation. Vazifeh et al. [2018] showed that the total fleet size increases by up to 10% for two competitors and up to 15% for three competitors. Séjournè et al. [2018] define a price of fragmentation and study it with the means of formal analysis and simulations assuming independent and sufficient vehicle supply.

AMoD Demand & Stochasticity of Customer Interaction: This paragraph can only provide a brief overview as a detailed review of descriptions, analysis, and models for possible travel behavioral changes in the presence of AMoD systems are out of the scope of this thesis. Many factors will influence future AMoD demand. It can also be expected that the attitude towards AVs in general and AMoD in specific will change over time. Prior knowledge from MoD systems is undoubtedly useful [Rayle et al., 2016; Bansal, Sinha, et al., 2020; Moody and Zhao, 2020; Tirachini, Chaniotakis, et al., 2020] as there are no large-scale pilots generating revealed preference data[3]. For the same reason, stated preference surveys are the method of choice to study various topics in the AMoD field, such as general AV acceptance [Abraham et al., 2017; Bansal and Kockelman, 2017; X. Dong et al., 2019; Kaltenhäuser et al., 2020; Nair and Bhat, 2021], pooling behavior [Gurumurthy and Kockelman, 2019; Y. Liu, Bansal, et al., 2019; Alonso-González et al., 2020; Hou et al., 2020], general mode choice and handling of MaaS platforms [Krueger et al., 2016; Yap et al., 2016; Frei et al., 2017; Caiati et al., 2019; Feneri et al., 2020], and more. From a modeling perspective, many studies applied frameworks with exogenous demand, i.e., the demand for the AMoD system was estimated beforehand and is deterministic and independent from the fleet operation. The generated demand can be based on taxi or MoD data [Santi et al., 2014; Dandl and Bogenberger, 2019; Dandl, M. Hyland, et al., 2020], estimated from a transportation model with assumed modal shift [Fagnant, Kockelman, and Bansal, 2015], or generated from integrated demand models (including population synthesis, trip generation, destination choice, and mode choice). The simulation

[3] Harb et al. [2018] provide an interesting approach to estimating induced demand on a small scale.

platforms SimMobility[4] and MatSim[5] are examples of the latter, which developed iterative frameworks, where mode choice is performed in a separate layer with the help of different logit models [Ben-Akiva and Bierlaire, 1999]. The mode choice and fleet simulations are iterated until an equilibrium is achieved [Marczuk et al., 2015; Sebastian Hörl, 2017]. A day-to-day adjustment process is applied to the traveler agents to allow exploration of new mode-choice alternatives and convergence towards (globally) stable and meaningful mode-choice decisions [Djavadian and J. Y. J. Chow, 2017; Basu et al., 2018; Sebastian Hörl et al., 2019]. Similarly, Kang et al. [2015] use a Hierarchical Bayesian choice model in a layer separate from the fleet simulation. Since the demand of these frameworks is not determined during the run-time of the actual fleet simulation, the demand can be viewed as predefined and deterministic from a fleet operator's perspective. Frameworks, where the acceptance of customers does not rely on time constraints (booking) or is only based on strict time constraints, have deterministic customer interactions. On the contrary, there are frameworks where customer interactions are based on fleet operational decisions in a probabilistic way. Atasoy et al. [2015], Dandl, Bogenberger, and Mahmassani [2019], and Wilkes et al. [2021] developed integrated mode-choice models, where the travelers base their decisions on real-time information. This information can contain fares, pick-up times, and drop-off times and is generated by the fleet operator based on the state of the fleet and traveler-specific information (origin, destination, request time). In these models, the AMoD demand is related to the alternative travel options. Another possibility is to assume sensitivities to certain aspects, e.g., fares and waiting times. J. J. Q. Yu et al. [2018], Turan et al. [2020], and Al-Kanj et al. [2020] built frameworks in which the amount of demand is probabilistic and sensitive to the AMoD fares set by an operator's dynamic pricing policy. Customers in the simulation environments of Engelhardt, Dandl, Bilali, et al. [2019] and Loeb and Kockelman [2019] accept the AMoD based on a waiting-time dependent probability. Cancellations of trips are another source of stochasticity that has recently been studied [X. Wang et al., 2019; J. Yu and M. F. Hyland, 2020].

[4] https://its.mit.edu/software/simmobility
[5] https://www.matsim.org/

2.3 AMoD Fleet Operation

Essentially, the operation of an AMoD fleet represents a vehicle routing problem, i.e., a fleet of vehicles receives assignments of routes and stops to perform specific tasks. There are various categories to distinguish different types of VRP. Psaraftis et al. [2016] describe a taxonomy with 11 criteria, most of which have been described in the service design. One of the most important criteria is the inflow of available information, denoted as the *type of problem* by Psaraftis et al. [2016] or *evolution of information* and *quality of information* by Pillac et al. [2013]. The evolution of information can be static or dynamic. In a static problem, all information is available at the time of decision-making, i.e., the time at which the routes and stops are assigned to vehicles. Contrarily, in a dynamic problem, information is revealed to the operator over time making re-assignments for the vehicles meaningful. Larsen [2000] describes various definitions for the degree of dynamism: (i) the share of customers that are revealed over time, (ii) the share of requests that have been revealed until a certain point in time, and (iii) the ratio of reaction time, which is defined as the time between a customer is revealed and a decision regarding this customer has to be made, and the planning horizon. Quality of information describes the stochasticity of information. If all information is deterministic, the problem is called deterministic. Otherwise, it is classified as a stochastic problem. Demand forecasts, which are different from exact future customer information as they predict demand in spatio-temporal aggregated form [Sayarshad and J. Y. J. Chow, 2016; Dandl, M. Hyland, et al., 2019], are an example of stochastic information in dynamic problems. The consideration of stochastic travel times is another example that typically include reliability as a quality measure of routing [Schilde et al., 2014; Filipovska and Mahmassani, 2020].

Pick-up and delivery problems constitute a subclass of vehicle routing problems, in which the operator has to fulfill requests to transport people or goods from an origin to a destination. If people are the customers and all information is available, the (static) problem is referred to as dial-a-ride problem (DaRP). In its original form, every transportation request from a set of customers asking for transport in a network must be served while satisfying certain time constraints. Cordeau and Laporte [2007], Molenbruch et al. [2017], and Ho et al. [2018] surveyed the literature for DaRP formulations and solution approaches. Moreover, the latter also summarized publicly available test instances[6] and the largest problem instances that could be solved exactly: 8 vehicles and 96 customers [Braekers et al., 2014; Gschwind and Irnich, 2015]. There are two-index (sequence of customer pick-up, and drop-off point) and a three-index (customer pick-up, drop-off and vehicle) formulations of the problem. Bongiovanni et al. [2019] extended two- and three-index DaRP problem formulations by including range constraints for electric vehicles and employed a branch-and-cut strategy to solve instances of up to 5 vehicles and 50 customers. Selecting the best of the combinatorial possibilities of sequencing pick-up and drop-offs is NP-hard. Even though these problem instances seem small, they are computationally expensive. In order to tackle larger VRP instances, meta-heuristic solution approaches were developed: simulated annealing (e.g., Chiang and Russell [1996]), genetic algorithms(e.g., Baker and Ayechew [2003]), large-neighborhood search(e.g., Ropke and Pisinger [2006], B. Li et al. [2016], and Syed, Kaltenhaeuser, et al. [2019]), Tabu search(e.g., Gendreau and

[6] See e.g. http://alpha.uhasselt.be/kris.braekers/

Potvin [2005] and Erdmann, Dandl, and Bogenberger [2019]), or hybrids of these algorithms(e.g., Vidal et al. [2012]).

In the following, it is assumed that the AMoD service is highly dynamic, i.e., customers use their smartphone to make immediate requests and expect fast decisions by the AMoD operator whether they are to be served. Actually, the restricted access to information and immediacy of response constrain the solution space severely compared to the static case. On the one hand, this leads to the solutions of a dynamic setting being worse than the static case solution; on the other hand, much larger systems can be addressed. The dynamic nature requires frameworks, in which the movements of the vehicles are simulated and the operator makes decisions based on the current system state [Regan et al., 1996; Horn, 2002; J. Yang et al., 2004; Jung and Jayakrishnan, 2014].

The following paragraphs and Table A5 (in Appendix D) divide the fleet operational decisions into a few categories and summarize the literature about the respective topics.

Fleet Sizing: One of the most critical factors regarding fleet efficiency is fleet size. If the fleet is very large compared to the demand, many vehicles will be idle producing costs without generating revenue. If all vehicles are utilized and there is unmet demand, the addition of vehicles could be profitable. In theory, AMoD operators could use two-sided platforms and surge pricing to increase an elastic fleet of borrowed privately owned AVs [Stocker and S. Shaheen, 2017; Nourinejad and Ramezani, 2019] similar to attracting drivers to be available during peak demand [Hall et al., 2015]. Since this thesis assumes that privately owned vehicles will play a minor role, AMoD providers also own their AVs and have a constant fleet size. As drivers are not needed, these vehicles will also be available for operations (unless they require maintenance). Therefore, driver scheduling and its impact on elastic fleet size do not have to be considered. Several approaches to fleet sizing are considered in literature. Analyzing demand and average trip duration gives an immediate estimation for a meaningful range of fleet sizes [Dandl, M. Hyland, et al., 2020]. Similarly, macroscopic approaches allow a fast estimation of impacts of various fleet sizes [Daganzo and Ouyang, 2019]. Alternatively, simulations with various explicit fleet sizes return the performance measures for the respective number of vehicles. For microscopic models, (warm-up) simulations allow the dynamic generation of vehicles if no other vehicle would be able to serve a customer; even though this procedure does not represent a real-time operation, it provides a well-sized fleet [Fagnant and Kockelman, 2014; T. D. Chen and Kockelman, 2016]. A commercial AMoD operator will typically optimize the fleet size with respect to profit, i.e., weight the generated revenue, the generated costs (by using the vehicles), and the fixed costs of vehicles [Dandl and Bogenberger, 2019; Sebastian Hörl et al., 2019]. These fixed costs typically consist of investment/leasing, mileage-independent deprecation, insurance, and parking fees [Bösch et al., 2017; Dandl and Bogenberger, 2019]. In a joint PT network design, other measures might be suitable for optimizing the fleet size [Pinto et al., 2019].

Customer Vehicle Assignment: The profit-generating and thereby main task of the fleet operator is the assignment of customers to its vehicles. Whereas an optimal assignment can clearly be defined for the static case, the dynamism of the problem does not guarantee that making the best assignments according to a control objective and the current state is better

than another assignment in the long run. Nevertheless, it makes sense to apply optimization techniques to assign customers to the best of the current knowledge [M. Hyland and Mahmassani, 2018a; Dandl and Bogenberger, 2018]. The dynamic re-optimization of the current state is denoted by dynamic state optimization in this thesis. For the ride-hailing case, M. Hyland and Mahmassani [2018a] studied the performance of different re-assignment policies: customers can be fixed to a vehicle after the initial assignment or be re-assigned to another vehicle. The other vehicle may currently be be idle or en-route to pick up or drop off another customer. In their problem formulation, en-route drop-off vehicles are assumed to first deliver the current on-board customer before driving for the assigned pick-up; vehicles that are re-assigned while en-route to pick-up do not pick up the previously assigned customer but are diverted directly to the newly assigned customer. They concluded that the more re-assignments are possible, the higher the optimization potential and the gain in overall fleet performance are. Heuristics are often used in the hailing case; some of them hardly affect the solution, e.g., when limiting the number of vehicles that are considered for each customer in the optimization [Dandl and Bogenberger, 2019], others severely limit the solution space. The most constraining but computationally fastest approach is the nearest neighbor policy, in which a new customer is assigned to the (idle) vehicle that can pick her up next. Maciejewski, Bischoff, and Nagel [2016] developed a widely-used insertion heuristic that is computationally as efficient as nearest-neighbor but searches for the next free customer in under-supply and performs much better in these situations. For pooling, the solution space grows very fast with the number of on-board and waiting customers. Therefore, many dynamic ride-pooling studies employ some insertion heuristic, in which a new request is included the existing route of a vehicle according to some heuristic rules [Jung, J. Y. Chow, et al., 2014; Fagnant and Kockelman, 2018; Ma et al., 2019]. Winter, Cats, G. Correia, et al. [2018] use a heuristic in which vehicles can accommodate other customers as long as these customers arrive within a certain dwell time at a stop. Alonso-Mora, Samaranayake, et al. [2017] developed the state of the art for pooling algorithms, which is a graph-based approach exploiting feasibility consideration to build a tractable optimization procedure. The idea behind the approach is that for hard time constraints (given by the service design), the number of routes that do not satisfy these constraints grows nearly the same as the number of possible combinations leaving only a few feasible routes for a larger number of customers. The authors mention that they were inspired by the approach of Santi et al. [2014] to compute the shareability of requests. In order to find these feasible routes efficiently, Alonso-Mora, Samaranayake, et al. [2017] proposed a multi-step approach, which builds feasible routes with an increasing number of requests. Later papers improved the performance for the optimal case [Y. Liu and Samaranayake, 2019; Engelhardt, Dandl, and Bogenberger, 2019], revisited the formulation for the two-request case [M. F. Hyland and Mahmassani, 2020], developed heuristics to limit the graph-building process [Simonetto et al., 2019; Engelhardt, Dandl, and Bogenberger, 2019], or changed the constraints of requests [Kucharski and Cats, 2020].

It has to be mentioned that the AMoD system modeling approach clearly affects the operational model. Studies assuming a station-based AMoD operation can apply a simple first-come first-serve queuing model at the station [M. Pavone et al., 2012; Spieser et al., 2016]. Studies assuming exact knowledge of future customers can build schedules to serve

these and transitions between the schedules [Boesch et al., 2016; Dandl, Grueber, et al., 2019; Wallar, Schwarting, et al., 2019]. Flow-based and macroscopic models typically do not consider an explicit assignment of a specific customer with a specific vehicle [Tachet et al., 2017; Salazar, Rossi, et al., 2018; Daganzo and Ouyang, 2019]. On the other spectrum of modeling, Dandl, Bogenberger, and Mahmassani [2019] took the expected computation time for an explicit decision based on optimization into consideration in the decision-making process.

Moreover, there are some non-myopic approaches, which consider additional aspects when making customer assignments. Alonso-Mora, Wallar, et al. [2017], Dandl, M. Hyland, et al. [2019], and Dandl, M. Hyland, et al. [2020] include variables and objective components for short-term forecasts in their assignment optimization problem formulation, Boewing et al. [2020] build a model-predictive control approach that also considers the scheduling of electric charging processes, and Al-Kanj et al. [2020] even consider dynamic pricing and repositioning in their customer assignment problem.

Dynamic Pricing: Dynamic pricing is a known concept in VRP [Figliozzi et al., 2007]. While dynamic pricing is often studied in the carsharing [Lippoldt et al., 2018; Lippoldt et al., 2019] and MoD hailing literature [L. Chen et al., 2015; Bimpikis et al., 2019; He and Shin, 2019; Nourinejad and Ramezani, 2019], it is less prominent in the studies about AMoD systems. In the current ride-hailing market, dynamic pricing is used to attract drivers to the system and balance demand and supply, both in the spatial and temporal dimensions. As the AVs are available 24/7, this supply-attraction mechanism is not necessary for AMoD systems. Nevertheless, there is still potential to improve the profit by dynamic pricing [Turan et al., 2020; Al-Kanj et al., 2020]. The relative gain depends on the fare and cost structures; dynamic pricing produced approximately 10-30 % gain in the respective case studies. Al-Kanj et al. [2020] use learning methods to estimate the probability of users accepting certain fares and the probability that a vehicle will be assigned to the origin zone of that user given a certain fare. Turan et al. [2020] assume the knowledge of a price sensitivity and use reinforcement learning to set the prices (as well as reposition and charge vehicles).

Repositioning: Moving vehicles pro-actively to balance demand and supply is an important feature of on-demand mobility systems. It is very prominent in carsharing, where a vehicle has a catchment area with a radius corresponding to the typical walking distance of customers. Staff is required to bring vehicles from cold to hot demand zones in order to increase vehicle utilization and profit [Weikl and Bogenberger, 2013; Weikl and Bogenberger, 2015]. In MoD and AMoD systems, the catchment area of a vehicle is larger as the distance, which a vehicle can drive in the time people are willing to wait, is larger than the typical walking distance. On the other hand, vehicle repositioning is much cheaper as the drivers are already in the vehicle in the MoD case, or AVs can drive themselves in the AMoD case. One crucial distinction to today's MoD hailing services is that a centrally controlled and optimized repositioning strategy can be much more efficient than a decentralized system, where each driver tries to maximize their own profit. Studies about ride-hailing systems report increased energy consumption due to empty travel of more than 40% compared to people using a private vehicle [Wenzel et al., 2019], whereas it is one of the objectives of a centrally controlled

fleet to minimize the empty mileage [M. Pavone et al., 2012; Dandl and Bogenberger, 2019; Dandl, M. Hyland, et al., 2019]. Due to the dynamism and stochasticity of demand, it is not intuitively clear how vehicles should be repositioned. Different algorithmic solutions have been developed in recent years. They usually involve the allocation of the operating area in disjoint zones, which are treated as separate virtual hubs or stations, for which the expected demand and vehicle supply are estimated. Most studies assume perfect demand and supply information, whereas Dandl, M. Hyland, et al. [2019] analyzed the impact of spatio-temporal aggregation on forecast quality and its impact on fleet performance. They found that the forecasts of taxi demand in Manhattan become more accurate the less precise they are, i.e., forecasts with lower spatial and temporal resolution showed smaller errors. Moreover, they found that a higher spatial resolution, despite lower accuracy, has positively impact fleet performance in their joint short-term and mid-term assignment and repositioning approach. By comparing with simulations based on *perfect forecasts*, which are 0-error forecasts by aggregating the actual trip data for given spatio-temporal resolution, they also noticed that the impact of forecast errors on fleet performance becomes more severe the higher the resolution. Another approach is to call the repositioning algorithm only once to rebalance the fleet overnight [Zhu et al., 2017]. However, most studies apply their respective repositioning algorithms periodically. A heuristic approach often utilized by studies is the block-balance algorithm developed by Fagnant and Kockelman [2014]. In their strategy, vehicles reposition to neighboring zones if their block-balance value, which is the difference between the share of vehicles and the share of expected demand, is larger than a certain minimum difference. Other heuristic approaches using local optimization are presented by Marczuk et al. [2015], Winter, Cats, Martens, et al. [2020], and Winter, Cats, Martens, et al. [2021], where vehicles are sent to the next depot or next free parking option. The widest-use optimization-based approach is the minimum transport problem, e.g., [M. Pavone et al., 2012; R. Zhang and M. Pavone, 2016; S. Hörl et al., 2019; Ruch, Horl, et al., 2018], in which the number of excess vehicles is distributed evenly among the zones and the objective is to minimize the traveled distance. The minimum transport problem is utilized with two demand-supply estimation processes: a pure feed-forward approach estimating arrival and departure rates and a feed-back approach based on currently open requests and vehicle movements. In a comparative study, the feed-forward approach showed more promising results [S. Hörl et al., 2019]. Other optimization-based approaches include a term for the imbalance in the objective function [Alonso-Mora, Samaranayake, et al., 2017; Wallar, van der Zee, et al., 2018; Dandl and Bogenberger, 2019], which are denoted as dynamic state optimization in Table A5. With hypothetical requests or demand-supply imbalances, it is also possible to integrate repositioning in the objective of the customer-assignment optimization [Alonso-Mora, Samaranayake, et al., 2017; Dandl, M. Hyland, et al., 2019]; such approaches are denoted by combined dynamic state optimization. In contrast to these integer optimization models, flow-based equilibrium models assume constant demand but are capable of integrating flow-based capacity or travel time models [Rossi et al., 2018; Salazar, Tsao, et al., 2019]. Previously described approaches consider constant demand or estimate the demand within a certain time horizon in one step. Dandl, M. Hyland, et al. [2020] apply a dual-horizon approach, where short-term forecasts are considered in the customer assignment and a longer time horizon is used for a separate repositioning problem.

Other model-predictive control approaches split the future time horizon into several steps and optimize the sum of the objective of all steps. The resulting problems were solved with various methods, e.g., heuristics [Albert et al., 2019], mixed-integer linear programming [R. Zhang and M. Pavone, 2016; Charkhgard et al., 2020], approximate dynamic programming [Al-Kanj et al., 2020], or reinforcement learning [Wen, Zhao, et al., 2017; Fluri et al., 2019]. These models for future impact can also contain other control aspects, such as charging [H. Zhang et al., 2019; Al-Kanj et al., 2020; Estandia et al., 2021].

Charging: The charging strategy determines when and where electric vehicles are sent to replenish their batteries. Operators can use public charging infrastructure or in their own depots. Studies typically assume that the AMoD operators do not have to share the charging infrastructure and can access charging units whenever their vehicles arrive at an empty charging unit. Strategies with different complexity have been developed in recent years. A trip-rejection criterion was developed by T. D. Chen and Kockelman [2016], which sends vehicles charging if they do not have sufficient energy to serve a customer request. This strategy is problematic in large operating areas, where a single request can send many vehicles to charge at the same time. Simple threshold models were applied in Dandl and Bogenberger [2019] and Hu and J. Dong [2020], where vehicles are sent to the nearest free charging station (after serving their current customers) when the state of charge of a vehicle drops below a certain threshold. Fehn et al. [2019] built a rule-based system that considers both state of charge and the current energy price. Furthermore, various model-predictive control approaches have been developed. Due to the curse of dimensionality, the charging strategy by Rick Zhang et al. [2016] practically only considered charging over a very short time horizon. For a less myopic approach, Iacobucci et al. [2019] added a second time horizon for charging, and Wenbo Zhang et al. [2019] and Estandia et al. [2021] used flow-based models. The models by Boewing et al. [2020] and Estandia et al. [2021] also consider the power network in their formulations. Dandl, Fehn, et al. [2020] built a macroscopic model to account for electricity prices, battery wear, and the influence of limited range and charging processes. They developed a model-predictive control approach to generating a macroscopic charging schedule for a whole day.

2.4 Regulation of AMoD Systems

In contrast to the vast amount of literature about the operational side, research on the regulatory side with AMoD systems seems somewhat limited. Hence, studies about AV regulation in general and MoD and taxi systems build the basis for the following paragraphs. Fraedrich et al. [2019] surveyed city administrations and identified potential fields of action. They stated that shared rather than private AVs are better for urban development strategies. Another finding was that there could be a mismatch between urban planning and industry priorities, and urban planners will require integrated transportation models considering AVs to make informed decisions. Cohen and Cavoli [2019] describe outcomes from a workshop about AV policy with stakeholders from academia, central and local governments, transport operators, NGOs, and consultants. They argue that a passive approach could result in a worse performance of

the transportation systems with respect to traffic flow and accessibility. They suggest a list of regulatory measures; once again, prioritizing shared over private mobility is one of them. They acknowledged that parking management and congestion pricing are possible. Lamotte et al. [2017] and Beheshtian et al. [2020] suggested reservation-based pricing schemes, in which AVs have to request space on the road. Similar to the energy market, capacity constraints can be considered and traffic flow improved due to the central monitoring and control. Additional to parking and congestion pricing, Narayanan et al. [2020a] mention intersection control, platoon size, variable speed limit, and dedicated lanes as other traffic control level policies that affect the traffic flow efficiency of AVs. W. E. Walker and Marchau [2017] suggest a regulatory framework for real-life implementation of policies: after identifying the objectives, constraints, and options, a basic policy should be set up. Since the introduction of AVs brings a lot of uncertainty, they suggest studying the robustness of the basic policy and reactions to vulnerabilities and developing opportunities in advance. A monitoring system should be set up to determine whether the underlying assumptions are still valid. Moreover, the monitoring allows a timely policy adaptation by triggering the respective actions.

It is worth noting that regulations are already in place for current taxi and MoD systems and have been studied in literature, e.g., by H. Yang et al. [2005]. Taxis are severely regularized; arguably even to the degree that the severe regulations paved the way for the major growth of MoD systems in comparison to taxi systems [Cetin and Deakin, 2019]. In Germany, there even exists a taxi regulation denoted 'Rückkehrpflicht', which means that taxis have to return to their original zone and are not allowed to serve customers on this return trip. This law was introduced to regulate competition between multiple taxi services but inherently causes the share of non-revenue miles to be about 50%, which obviously is inefficient and bad for the environment. As MoD systems started to grow, state and city administrations introduced some regulations, e.g., background checks of drivers, driver's license requirements, vehicle requirements, and the requirement for MoD providers to share data with administrations [Beer et al., 2017]. More recently, the administration of New York City limited the number of MoD vehicles [Bellon, 2019] and approved plans to introduce congestion pricing in Manhattan [Campbell, 2019; Conger, 2020]. Such measures were studied by S. Li et al. [2019] and K. Zhang and Nie [2019]; however, the frameworks are not directly transferable to the AMoD case since the two-sided MoD market structure and minimum wages are key components of their respective methodologies.

Some studies considered transportation systems with AMoD services and some kind of regulatory measure in place. Simoni et al. [2019] analyzed the impact of different congestion pricing schemes in the presence of AMoD. They showed that marginal-cost and travel-time dependent pricing schemes could reduce total VKT and improve social welfare, which they defined as the sum of toll revenues and all traveler's utility. Basu et al. [2018] studied the case of removing the existing mass transit in a virtual city reflecting Singapore, which resulted in high levels of congestion. In a similar study, Oke, Akkinepally, et al. [2020] analyzed the removal of mass transit and the integration of AMoD as a feeder system for car-dependent prototype cities. Under the assumption that the hailing-to-pooling ratio is similar to today's MoD systems, they observed an increase in VKT and congestion. Recognizing that a complete replacement of mass transit is not meaningful, Becker et al. [2020] removed only the least efficient PT lines. This method resulted in reductions of emissions and positive social welfare

results, even though they did not consider pooling of customers. They also mentioned that the full costs rather than only the operating costs of private vehicles should be considered for a long-term view.

2.5 Conclusion

Research about AMoD systems has been very active in recent years. As a result, many service designs and operational strategies have been developed and evaluated with various modeling techniques. However, improvements for the dynamic and stochastic fleet operational problems remain possible, especially when considering customer assignment, dynamic pricing, and repositioning all at once and for different service designs.

A smaller amount of studies focused on the regulation of transportation systems with AMoD services. Some regulatory measures are transferable from real-world regulation of MoD systems and studies thereof: a limit on fleet size, congestion pricing, and parking management certainly are applicable for steering demand and supply in a transportation system with AMoD services. Moreover, shared AVs are to be prioritized over private AVs and pooling will be required to increase the average occupancy and thereby improve system VKT and the general traffic situation.

Research gaps can be identified when it comes to the interplay between operator and regulator. The few studies testing regulatory measures did not consider that the operator will react to these measures, i.e., optimize its system in the new setting. Additionally, the regulatory actions mainly were of qualitative nature and studied one measure at a time. However, if fleet size limit, congestion pricing, and park costs are to be adapted, regulators need to know which quantitative values they should set. They also have to consider the interplay of multiple regulatory measures acting at once.

The main contributions of this thesis are

- the creation of further insight into operational aspects of AMoD systems.
- the development of a new optimization-simulation framework that considers these operational insights to regulate AMoD systems.

The envisioned framework needs to contain AMoD models that are sensitive to regulatory measures. There is a trade-off between granularity and computational complexity. Even though macroscopic models are often better generalizable, they typically assume a high degree of homogeneity. In addition, modeling the sensitivity to the regulatory measures is challenging or sometimes impossible. Agent-based models can accommodate more heterogeneous data input but require more computational effort. Since resources are limited, it is necessary to focus on certain operational variables before going to the regulatory part of the thesis.

Chapter 3

AMoD Service Design and Fleet Operation

This chapter assumes a laisser-faire scenario, where no regulatory measures are in place. Instead a system, with an AMoD operator and travelers are studied from the operator's view. In the first section, certain service design models will be selected. For these service designs, the mathematical problem will be formulated and the operator's actions/strategies will be introduced. These actions are inherently related to different time scales, which allow a divide-and-conquer approach to tackling the different possible actions in the subsequent sections. A case study will be used to gain further insights on the impacts of these actions. Finally, key learnings are summarized in a brief conclusion of the chapter. During the course of this chapter, mathematical notation is introduced, which enables a brief and precise formulation. A summary of the mathematical notation is provided in Appendix A.

3.1 Problem Description

This thesis studies an AMoD service operating a homogeneous fleet of vehicles $v \in V$ in a street network $G = (N, E)$, in which (i) a set of potential users R make app requests to the operator, (ii) the operator has to respond immediately with an offer or a rejection, and (iii) the users accept or reject an offer, if available. The resulting operational problem is dynamic as requests are revealed over time. In most scenarios, users expect to be served as soon as possible; however, a strategy for user reservations is developed and tested in the case study. Users communicate their time constraints with the request and do not accept offers that do not satisfy these constraints; moreover, users are modeled price sensitive in the scenarios testing pricing strategies. The offers, which are communicated from the operator to the users, have to be created by a strategy that guarantees the level of service[1]. Therefore, the fleet control strategies has to avoid to confirm a service but then not deliver the promised service.

The fare structure of the AMoD service is assumed to have a constant base fare (f^C) and a distance-dependent (f^D) component. For simplicity, it is assumed that the constant cost structure of the AMoD service scales with fleet size: all time-based costs (e.g. staff, platform costs, marketing, vehicle investment/lease, insurance) are distributed per vehicle on a daily basis and denoted by fixed vehicle costs (c^F). Energy costs, vehicle depreciation, battery wear

[1]The alternative would be a game, in which both operator and users speculate about the offered and actual level-of-service, which is out of scope of this thesis.

are assumed constant per driven distance and their total is denoted by variable vehicle costs (c^D). It is assumed that vehicle batteries have sufficient range in order to schedule charging in low-demand times and charging processes thereby do not restrict the fleet operation. Finally, it is assumed that street parking is allowed for AMoD vehicles; a scenario where vehicles have to return to hubs instead is tested in the case study to quantify the benefits of allowing street parking.

A request $r \in R$ is specified by a request time τ_r, a pick-up location $x_r^p \in N^a$, and a drop-off location $x_r^d \in N^a$, where $N^a \subset N$ is the set of access points defined by the operator. For the standard case, a maximum waiting time τ_r^w and a maximum driving time τ_r^d are also communicated. These values constrain the level of service a user can expect: τ_r^w determines the latest pick-up time and τ_r^d the maximum in-vehicle duration, which is relevant for pooling. In the reservation scenario, the operator offers a time-window in which a user will be picked up. This time window consists of an earliest pick-up time τ_r^e and a latest pick-up time τ_r^l. In this thesis, the latest pick-up time is given by $\tau_r^l = \tau_r^e + \tau_r^w$. Furthermore, this thesis assumes a quasi door-to-door service, i.e., requests are to be picked up at and dropped off at x_r^p and x_r^d, respectively. The operator can reject the request or return an offer, which contains an expected waiting time t_r^w, an expected driving time t_r^d, and a fare f_r, which is computed by

$$f_r = \max\left(f^D \cdot d_r, f^B\right) \tag{3.1}$$

where d_r is the distance of the fastest route between x_r^p and x_r^d, f^B is a minimum base fare and f^D is a direct-route distance-dependent fare component. This formulation ensures that users do not have to pay for any detour distance. Moreover, f^B is introduced to dissuade users from making very short AMoD trips. The number of passengers ρ_r that are part of a request has to be smaller than the seat capacity ρ_v of vehicles. For the pooling case, ρ_r determines how many of a vehicles available seats ρ_v will be taken by one request. Group discounts could be introduced easily but are not considered in this thesis for the sake of simplicity and tractability.

The operator has control over the fleet of vehicles V. A vehicle $v \in V$ can pick-up and drop-off users at their respective locations. The boarding and disembarking processes are assumed to last for τ^b. The operator assigns a vehicle plan ξ_v to each vehicle v. These plans contain stops X_s (with stop index $s \in \{1, 2, ..\}$) to pick-up and drop-off users and drive to certain destinations and wait for further instructions, i.e., repositioning. A stop $X = (x, t^s, t^l, t^e, R^+, R^-, lock)$ is defined by a network position $x \in N$, a stop duration t^s, possibly a latest arrival time t^l and an earliest departure time t^e, the sets of boarding and disembarking requests denoted by R^+ and R^-, and a flag $lock$ indicating a locked status. A stop with $lock = True$ cannot be changed anymore. It is set to *False* for short-term decisions but becomes relevant for the mid-term decisions. The stop duration is given by the boarding time t^b of AMoD system users. It is assumed that users disembark and board sequentially in a hailing operation, whereas boarding processes of multiple users at the same location x can be combined in a pooling service: all passengers of one stop can board and disembark within t^b. The latest arrival time at the stop (t^l) can be derived from the time constraints of the requests (both τ_r^w and τ_r^d). Without reservations, t^e of subsequent stops can be ignored as vehicles should always pick-up other requests of the route as soon as possible and hence leave a stop as soon as all boarding processes are finished. Obviously, a vehicle assigned to serve a reservation request r should not leave before the user had a chance to join the vehicle, i.e., not before $t^e = \tau_r^e + t_s$.

It is assumed that vehicles follow the fastest route ϕ between two subsequent stops. In a hailing operation, the drop-off stop of each request must immediately follow the pick-up stop. Contrarily, vehicle plans of a pooling operation have to satisfy the conditions that the pick-up stop comes before the drop-off stop but not necessarily immediately for each user. Furthermore, the vehicle route has to satisfy the condition that the vehicle cannot accommodate more than ρ_v passengers on board at a time.

Let T be a time horizon, for which operational decisions should be evaluated, which is typically in the range of a day. Moreover, let R be the set of all possible requests within that time horizon. The global objectives of the operator are to maximize market share and profit. The market share term penalizes not serving a request with a cost c^- per user; this term can also be interpreted as an opportunity cost as users who are not satisfied with the service on one day, might not even make a request in the following days. Therefore, the global operator objective P reads

$$P = \sum_{r \in R^s} f_r - \sum_{r \in R^l} c^- - \sum_{v \in V} \left(c^F + c^D \cdot d_v \right) \tag{3.2}$$

where R^s and R^l are the set of served users and the set of not served users that left the system, respectively. Moreover, d_v denotes the total distance that vehicle v drove throughout the time horizon. For meaningful fare and cost structures ($f^D > c^D$, $f^B \geq 0$), both objectives are in line[2]. Hence, it is in the interest of the operator to serve as many requests as possible. In this case, additional explicit level-of-service measures (for additional waiting and detour time) are not necessary in the global objective, because users reject offers not satisfying their time-constraints.

c^- can be interpreted both as assignment reward or penalty for not served requests. This can be shown using an alternative objective P'

$$P' = \sum_{r \in R^s} \left(f_r + c^- \right) - \sum_{v \in V} \left(c^F + c^D \cdot d_v \right) \tag{3.3}$$

The optimization with equations (3.2) and (3.3) as objectives is equivalent for a constant set of requests R, i.e., $R^s = R \setminus R^l$, because $P = P' + Const.$ In the introduction phase, large values of c^- make sense in order to expand. This thesis focuses on AMoD systems, which assume a long-term view with demand and supply in equilibrium. In this phase, the AMoD operator is more concerned with making profits than expanding its market share and the uncertain parameter c^- can be dropped. With $c^- = 0$, equations (3.2) and (3.3) are equal and denote the global operator objective P, which can be interpreted as the operator profit. c^- could be re-introduced by substituting $f_r \rightarrow f_r + c^-$ in any subsequent equation.

With this objective in mind, the AMoD provider has following set of actions A:

- determine the fleet size and operating area
- develop a fare system (standard fares and dynamical adaption)
- determine imbalance of demand and supply

[2]To be more precise, the fare has to be higher then the costs related to the vehicle driving to the user and serving the user.

- distribute vehicles in the operating area
- receive app requests; determine level-of-service; make offers
- assign a confirmed request to a vehicle
- build a vehicle plan

The actions A generally depend on the state S of the AMoD system. This can be modeled as a Markov Decision Process. Whenever the state changes, the operator might think about taking new actions (event-based approach).

In this formulation, actions can be assigned a (state-dependent) reward $\Psi = \Psi(S, A)$ such that the overall reward matches the global objective:

- the fleet size is penalized with the vehicle fixed costs c^F
- a confirmed user (as a reaction to a good offer action) is rewarded with the respective fare f_r
- the assigned vehicle plan ξ_v determines the distance d_{ξ_v} to pick up and drop off these users and reposition vehicles in the operating area; $c^D \cdot d_{\xi_v}$ are the corresponding costs (for driving)

As time is a state variable, a time-based formulation is possible, where all state changes between two time steps are accumulated and $S = S(t)$. Therefore, all actions performed at one time step can be collected $A = A(S(t))$. Moreover, the system state after performing all these actions, e.g. the assignment of a new vehicle plan, is changed and denoted by post-state $S^P(t)$. The state transition from time t_i to $t_{i+1} = t_i + \Delta t$ is depending on the post state and changes in the exogenous state variables $s(t_{i+1})$, which are independent of any actions the operator takes, e.g. requests that are revealed to the operator at time t_{i+1}. The state transition operator is denoted by Θ, i.e., $S(t_{i+1}) = \Theta(S^P(t_i), s(t_{i+1})) = \Theta(S(t_i), A(t_i), s(t_{i+1}))$. For a briefer notation, $S_i = S(t_i)$, $S_i^P = S^P(t_i)$, $s_i = s(t_i)$ and $A_i = A(t_i)$ are defined. The state $S(t_i)$ contains following attributes:

- the current position x_v for each vehicle v
- the current on-board requests R_v^o for each vehicle v
- the currently assigned vehicle plan ξ_v for each vehicle v
- the currently active requests R^a, which are those requests that have been revealed to the operator and have not yet been served or left the system
- the current state of the network
- stochastic information about future demand Λ and the network

The stochastic information Λ can be available in various forms and will be discussed further in section 3.4.1. For simplicity, the state of the network is assumed exogenous for this part of the thesis and sufficiently described by the travel times t_e for all edges $e \in E$.

The operator control problem can be summarized by

$$\min_{\boldsymbol{A}} P = \sum_{i=1}^{N^T} \Psi(S_i, A_i) \tag{3.4a}$$

$$\text{s.t. } S_{i+1} = \Theta(S_i, A_i, s_{i+1}) \qquad \forall i \in 1, ...N^T \tag{3.4b}$$

where $\mathbf{A} = \{A_i\}$ denotes the set of all actions taken at all time steps.

In theory, the Bellman equation would determine the set of optimal actions A_i at each time step t_i ($\forall i \in 1, ...N^T$):

$$A_i^* = \arg\max_{A_i} \left(\Psi(S_i, A_i) + \mathbb{E}\left[\sum_{t=i+1}^{N^T} \gamma^{t-i} \Psi(S_t, A_t) \right] \right) \tag{3.5a}$$

$$\text{s.t. } S_{i+1} = \Theta(S_i, A_i, s_{i+1}) \tag{3.5b}$$

where N^T is the number of time steps until the end of the time horizon to be evaluated and $\gamma \in]0, 1]$ a discount factor weighting future rewards against immediate rewards. The expectancy value is required if stochastic information is available and can be computed with Monte-Carlo simulations. In theory, this problem could be solved by dynamic programming. For problems of very small scale, the expectancy value could be computed by drawing requests from stochastic distributions and computing the rewards and impacts of the decisions to be made.

3.2 General Solution Approach: Separation of Time Scales

In practice, the action space is too large to apply dynamic programming to the equation system (3.5). General approaches to approximate such problem are available [Bertsekas, 2005; Powell, 2011]. However, the consideration of different time scales of the actions allows to divide the general problem into smaller and easier to tackle problems.

Similar to Atac et al. [2019], three time scales and classes of decision-making/actions are defined and illustrated in Figure 3.1. The outcomes of planning decisions remain constant for a long time. It does not make sense that the AMoD provider recomputes its investments into AVs based on single requests. It can be expected that users of an AMoD system want to see stability in operating area and approximate fare range in order to trust such a system. Hence, this thesis assumes that these decisions remain constant over the evaluation time horizon T (scale of a day). On the other end of the scale, short-term operational actions like answering app requests and request-vehicle assignments have to be performed very frequently, namely in the scale of seconds. These request-vehicle assignments should at least have enough foresight to consider different possibilities to reach a request before the maximum waiting time τ^w.

Hence, these short-term decisions naturally consider a time horizon in the range of a few minutes. Finally, forecasts of the stochastic demand for time horizons in the range of minutes to hours can improve the fleet control by dynamically distributing vehicles to balance expected demand and supply and setting dynamic fares in order to reduce expected demand to match supply if necessary. As these forecasts are also stable and do not change every second, these tactical (mid-term) decisions can be made in the minute-to-hour scale.

With this separation of time scales and the division of actions into these scales, the Bellman equation is effectively approximated by following equations:

$$A_i^{S,*} = \arg\max_{A_i^S} \tilde{\Psi}(S_i, A_i^S) \qquad \forall i \in 1, ...N^T \tag{3.6a}$$

$$A_i^{M,*} = \arg\max_{A_i^M} \left[\sum_{t=i+1}^{N^{T^M}} \bar{\Psi}(S_t, A_i^M; A_i^{S,*}) \right] \qquad \forall i : t_i \% \Delta t^M = 0 \tag{3.6b}$$

$$A_0^{L,*} = \arg\max_{A_0^L} \mathbb{E} \left[\sum_{t=0}^{N^T} \Psi(S_t, A_0^L; A_i^{M,*}; A_i^{S,*}) \right] \tag{3.6c}$$

$$\text{s.t. } S_{i+1} = \Theta(S_i, A_i, s_{i+1}) \qquad \forall i \in 1, ...N^T \tag{3.6d}$$

where A_i^S, A_i^M, A_0^L denote the short-term, mid-term, and long-term actions, respectively. Moreover $\tilde{\Psi}$ and $\bar{\Psi}$ are reward-function approximations to estimate the impact of short-term and mid-term decisions at time t_i on rewards within the respective time horizons. In comparison to the original formulation, the expectancy value and the discount factor γ are absorbed in these approximations. Equation (3.6a) states that short-term decisions are based on the current system state. Furthermore, equation (3.6b) states that the algorithm to determine repositioning is only performed every Δt^M. $\%$ is the modulo operator and obviously for this to work, Δt^M has to be a multiple of Δt. This equation assumes that the main impact of mid-term decisions at time t_i is within the mid-term time horizon $[t_i, t_i + T^M]$. The idea behind this approximation is that vehicles will be moved based on other decisions after t_i so that the effect of any single decision at t_i on decisions and rewards becomes smaller and smaller over time. This should not mean that it cannot influence the fleet performance at later times; an expected vehicle deficit in 3 hours has to be taken care of at some time, but the difference of doing it now or in 15 minutes is likely rather small; on the other hand, if the deficit is expected to occur in 15 minutes, it is important to reposition vehicles right away. Finaly, equation (3.6c) states that long-term decisions have to be made (i.e., planned) before the system can serve any users, i.e., time step 0. Finally, equation (3.6d) ensures that the system dynamics remain the same as in the original problem.

Strategies to address these sub-problems efficiently are the topic in the remainder of this chapter. First, the most immediate and essential operational problem, i.e., the interaction with users, is addressed; then, mid-term decisions are explored. As mentioned, the effect of long-term decisions can be assumed constant over the time horizon T. The section about these long-term decisions briefly discusses how an operator can address the respective variables.

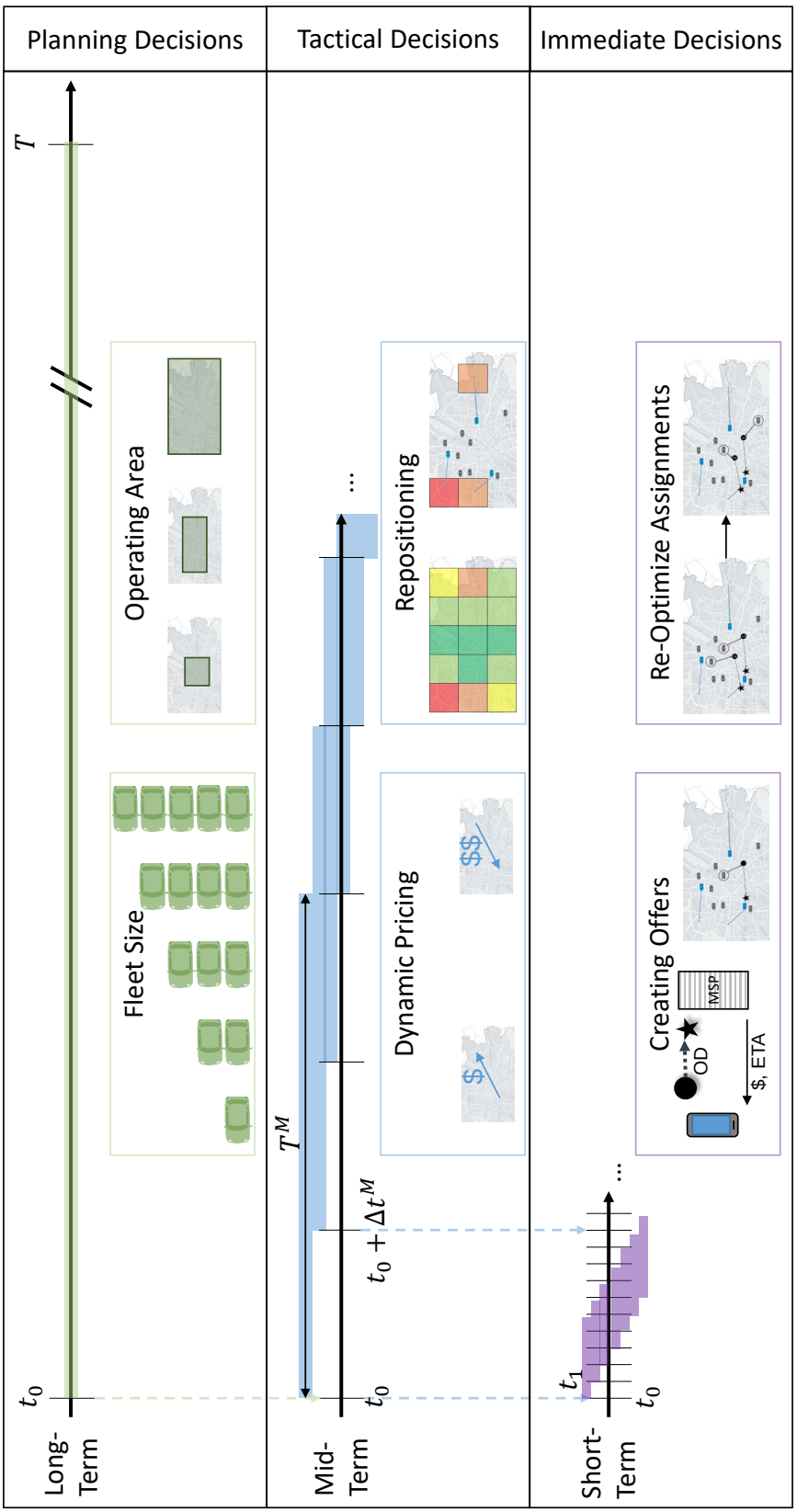


Figure 3.1: Time scales of AMoD operator actions. Long-term decisions like fleet size and operating area do not change within the evaluation time horizon (typically in the range of a single or multiple days), mid-term decisions like dynamic pricing and repositioning apply a rolling horizon approach, where decisions are based on a medium time horizon (typically in range of minutes to hours) with low frequency, and short-term decisions are performed very frequently.

3.3 Immediate Short-Term Decisions

This thesis assumes that the operator receives users' (app) requests and responds to them sequentially. The operator algorithm to respond should be fast such that users receive the offer practically instantaneously. As mentioned in the global problem formulation (section 3.1), the offers should be realistic. This is achieved in this thesis by making an actual request-vehicle assignment for an incoming request. The global performance can be enhanced by periodic re-optimization of request-vehicle assignments every Δt^S (in the range of seconds up to a minute).

Furthermore, it is assumed that users have already planned their decision-making process before making the request, i.e., they know which time and monetary constraints they will accept or reject and therefore can respond to an offer instantaneously[3]. With the requirements of instantaneous communication, requests that are revealed at time t either leave the system or become assigned to a vehicle v. There are no remaining open requests. The set of active requests (R^a) contains the disjoint subsets of requests waiting for pick-up ($R^w = \bigcup_v R_v^w$) and the on-board requests ($R^o = \bigcup_v R_v^o$).

Finally, it is assumed that the operator will not reject any request with start and end location in the operating area if it is able to serve this request within the given time constraints. However, the operator can increase fares if a vehicle shortage makes it unlikely that the operator can serve all incoming requests (see dynamic pricing in section 3.4.3 and its impacts in section 3.6).

In theory, it would be possible to assign only idle vehicles to user requests, at least in the hailing case. However, as shown by Dandl and Bogenberger [2018] and M. Hyland and Mahmassani [2018a], request-vehicle assignment algorithms with consideration of en-route vehicles severely outperform algorithms assigning only idle vehicles. In the pooling case, it is even inevitable to forecast routes in order to pool users together. Hence, the short-term decisions should have some foresight, at least in the range of the typical maximum waiting time (hailing) or average trip duration (pooling) of requests. The task of the short-term actions is the creation of vehicle plans that contain the stops to pick up and drop off the active requests. In order to check whether the plan ξ_v of vehicle v satisfies all time constraints of requests R_{ξ_v} considered in this plan, the expected arrival and departure times can be forecast for each stop. Ignoring reservation requests for the moment, the vehicle will stop at its planned stops for the boarding duration and travel along the fastest paths between stops (denoted by $\phi \in \xi_v$). A vehicle plan ξ_v is defined as feasible if the following conditions are satisfied:

$$x_r^p < x_r^d \qquad \forall r \in R_{\xi_v} \tag{3.7a}$$

$$\sum_{r \in R_v^\phi} \rho_r \leq \rho_v \qquad \forall \phi \in \xi_v \tag{3.7b}$$

$$t_r^w \leq \tau_r^w \qquad \forall r \in R_{\xi_v} \tag{3.7c}$$

$$t_r^d \leq \tau_r^d \qquad \forall r \in R_{\xi_v} \tag{3.7d}$$

[3] Even though the instant decision-making assumption is not realistic for all users of an AMoD system, it might be true for a majority as these service design parameters might be advertised openly. Moreover, this user-operator interaction model will turn out to be very advantageous in chapter 4.

The notation $x_i < x_j$ in equation (3.7a) means that a stop (for pick-up or drop-off) at location x_i has to be conducted before a stop at location x_j. Therefore, equation (3.7a) guarantees that each request has to be picked up before it is dropped off. With R_v^ϕ denoting the number of on-board requests on route $\phi \in \xi_v$, equation (3.7b) ensures that a vehicle plan cannot include a route ϕ with more than ρ_v passengers on board. Equations (3.7c) and (3.7d) check the waiting and detour time constraints of all requests in the plan, respectively.

Equations (3.7) have to be tested when a vehicle plan is built from scratch or a new request's pick-up and drop-off stop are inserted into an existing vehicle plan.

In order to select between different vehicle plans, the operator defines a (short-term) control objective F^S, which represents the short-term reward function approximation $\tilde{\Psi}$. F^S should be a function that approximates the impact of a request-vehicle assignment (and its respective vehicle plan) on the global objective $\Psi(S_i, A_i^S)$. It can contain a single objective component or be a multi-objective function. The scale of F^S as a whole can be chosen arbitrarily (i.e., F^S does not have to reflect the actual monetary reward) since for any $c_1 \in \mathbb{R}^+, c_2 \in \mathbb{R}$, following equation holds

$$\arg\max_A \left(c_1 \cdot F^S(A) + c_2\right) = \arg\max_A F^S(A) = \arg\min_A \left(-F^S(A)\right) \tag{3.8}$$

meaning that the solutions of the corresponding optimization problems are identical. However, if the optimization is multi-objective, then the scaling of single objectives is significant.

In the two following subsections, the two short-term operating tasks to create offers and re-optimize the operation are studied for hailing and pooling, respectively. In theory, the pooling methodology would also work for the hailing case but the simplifications for a hailing service improve computational efficiency.

3.3.1 Creation of Offers and Re-Optimization of Vehicle Plans (Hailing)

For the hailing case, in which at most one user is in a vehicle at a time, the drop-off stop of a request has to immediately follow its pick-up stop. When considering the insertion of a new request's pick-up and drop-off stop, only the waiting time conditions (3.7c) have to be checked while all other conditions (3.7) are satisfied automatically with this rule. As in M. Hyland and Mahmassani [2018a] and Dandl and Bogenberger [2018], this thesis assumes that it is sufficient to consider at most one pick-up process per vehicle plan. This assumption is valid if it is not possible to pick-up and serve a request r within the maximum waiting time of another request r', i.e., $t_r^w + t[x_r^p \to x_r^d] > \tau_{r'}^w$ is satisfied for all requests r, r'. This assumption is quite realistic for an AMoD hailing service that should not be used for very short trips and offers short waiting times of just a few minutes. If this assumption is not justified, the more general, but less efficient method described in section 3.3.2 should be applied. The method illustrated in the following reduces the creation of vehicle plans to an assignment problem (r, v) that aims to maximize the control objective function F^S. Once a vehicle v is selected for a newly revealed or waiting requests r, the plan stops $X_r^p = (x_r^p, t^b, \tau_r + \tau_r^w, \tau_r)$ and $X_r^d = (x_r^d, t^b, \tau_r + \tau_r^w + t[x_r^p \to x_r^d], \tau_r)$ are appended to the vehicle's previous plan.

The concept of spatio-temporal vehicle availability is applied to find efficiently the set of vehicles V_r that can reach a request r within its maximum wait time constraint. A backwards-

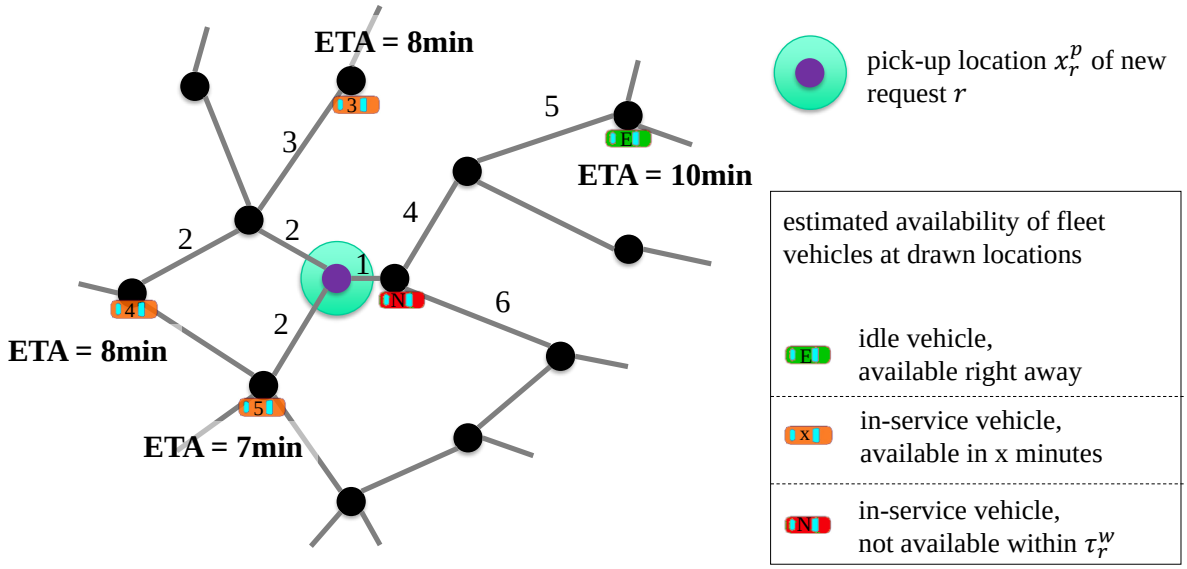


Figure 3.2: Search for available vehicles and estimated time to arrival (ETA) in the hailing case with maximum waiting time $\tau_r^w = 10$ minutes. Figure adapted from Dandl and Bogenberger [2019].

directed Dijkstra search around the request's pick-up location x_r^p with time as cost function and search radius τ_r^w can be utilized to determine the set of nodes N_r, from where a vehicle would have to start at time τ_r to reach r in time. In the second step, the availability of vehicles $v \in V$ at these nodes is considered. The set of vehicles V_r, which are feasible to serve request r, consists of vehicles that are either idle at position $x_v \in N_r$ or still in-service for another request r', in which case $x_v = x_{r'}^d \in N_r$. Without pooled rides, a vehicle with on-board passenger r' has to drop-off r' at $x_{r'}^d$ before being able to drive to x_r^p. As illustrated in Figure 3.2, such a vehicle v will be available at $x_v = x_{r'}^d$ after dropping of r' at time $t_v \geq \tau_r$. If the operator assigned vehicle v to request r at the current time t, the estimated time to arrival $t_{rv} = (t_v - t) + t[x_v \to x_r^p]$ would depend on the pick-up route time $t[x_v \to x_r^p]$ and the time of availability t_v. Hereby, the time of availability is the current time ($t_v = t$) for idle vehicles and in the future ($t_v > t$) for non-idle vehicles. The driving costs incurred by an assignment (r, v) are determined by $d_{rv} = d[x_v \to x_r^p]$. The pseudo-code of the search algorithm can be found in Appendix B.2.

The selection of a request-vehicle assignment (r, v) is determined by the control objective function F^S. It is important to note that an optimal control function cannot be derived directly due to the dynamic nature of the problem described by equations (3.4), i.e., the repeated application of the optimization procedure does not guarantee optimality with respect to the global objective. Instead, heuristic approaches have to be applied to choose a meaningful control function formulation.

Obviously, the empty distance to pick up a new request, i.e., $d_{rv} = d[x_v \to x_r^p]$, generates costs c^D and should be minimized. The costs to drive from x_r^p to x_r^d could be included as well but they would not change the result as they are the same for each vehicle. It can make sense to include a term to minimize user waiting times by minimizing the estimated time to

arrival t_{rv}. Choosing a vehicle with the lowest t_{rv} increases user satisfaction as a user r is picked up and dropped off earlier. It can also be beneficial to the system because the vehicle v that serves r becomes available as soon as possible. This thesis does not consider strategic rejections, which weight immediate profit against an expected revenue that the vehicle could make by remaining available and serve a later request; the goal of the operator in this thesis is to serve incoming customer requests if possible. This thesis models operators, which aim to minimize the following control objective function for the assignment of a hailing-request r to vehicle v

$$F_{rv}^{S} = c^{D} \cdot d_{rv} + c^{VOT} \cdot t_{rv} \tag{3.9}$$

Here, c^{D} and c^{VOT} are constant distance cost and value-of-time cost coefficients, respectively. A real AMoD operator will likely try to incorporate various other terms and vary the two coefficients over time and evaluate the performance of the respective control function in simulations.

Creation of Offers (Hailing): Whenever a request r is revealed to the operator, a local optimization is performed to determine the vehicle v this request would be assigned to. This local optimization is also denoted by insertion heuristic, which in the hailing case is a modified nearest-neighbor policy.

As a first step, the set of vehicles V_r is determined via the backwards-directed Dijkstra method. Then, the operator picks the vehicle v^* for request r that fulfills

$$v^* = \underset{v \in V_r}{\arg\min}\, F_{rv}^{S} \tag{3.10}$$

Note that v^* would also be the best solution for request r if the search would include all vehicles in V.

If $V_r = \emptyset$, the operator is not able to serve the user and has to reject the request. Otherwise, the vehicle plan of v^* is extended to include pick-up and drop-off stops at x_r^p and x_r^d, respectively. The operator already knows the estimated waiting time t_d^w, can estimate the in-vehicle time $t_r^d = d[x_r^p \rightarrow x_r^d]$, and compute the fare in order to create an offer to r. If r accepts the offer, the stops for request r are appended to the vehicle plan of v^* and are considered for the future availability of this vehicle. If v was idle, it starts driving towards x_r^p. If it is en-route to serve another user, it will continue to do so and drive to x_r^p after dropping off the current request. If r declines the offer, the original vehicle plan of v^* is restored, i.e., the stops to pick-up and drop-off r are removed.

Re-Optimization of Vehicle Plans (Hailing): In order to improve the assignments made by sequential local optimizations for each new request, the request-vehicle assignments are globally optimized periodically. Every Δt^S, a bipartite linear assignment problem is created and solved, which considers all possible assignments of requests $r \in R^w$ waiting for pick-up and vehicles $v \in V$. The requests that are on-board of a vehicle do not have to be considered in the assignment problem. With the control objective function F_{rv} defined in equation (3.9),

the optimization problem reads:

$$\min_{\boldsymbol{\beta}} \sum_{r \in R^w} \sum_{v \in V} \beta_{rv} \cdot F^S_{rv} \tag{3.11a}$$

$$\text{s.t.} \sum_{r \in R^w} \beta_{rv} \leq 1 \qquad \forall v \in V \tag{3.11b}$$

$$\sum_{v \in V} \beta_{rv} = 1 \qquad \forall r \in R^w \tag{3.11c}$$

$$(t_{rv} - (\tau^w_r - (t_i - \tau_r))) \cdot \beta_{rv} \leq 0 \qquad \forall r \in R^w \;\; \forall v \in V \tag{3.11d}$$

$$\beta_{rv} \in \{0, 1\} \qquad \forall r \in R^w \;\; \forall v \in V \tag{3.11e}$$

β_{rv} denotes a binary variable that is one, if an assignment between request r and vehicle v should be made. Equation (3.11b) ensures that each vehicle has at most one waiting request assigned, equation (3.11c) guarantees that each request, which was previously confirmed and assigned, is assigned to a vehicle. Moreover, equation (3.11d) is necessary to only make assignments that satisfy the time constraints of requests; either the estimated time to arrival t_{rv} is smaller than the remainder of the maximum waiting time or $\beta_{rv} = 0$.

Instead of creating variables and computing the route costs d_{rv} and t_{rv} for each request-vehicle pair, the number of variables can be reduced to speed up the optimization process. The backwards-directed Dijkstra method can be used for each request $r \in R^w$ to determine the set V_r that can satisfy equation (3.11d). A reduced bipartite graph can be created, in which edges between r and v only exist, if $v \in V_r$ for each r. This procedure allows the definition of R^w_v as the set of requests in R^w that are connected to a vehicle v. The reduced optimization problem then reads:

$$\min_{\boldsymbol{\beta}} \sum_{r \in R^w} \sum_{v \in V_r} \beta_{rv} \cdot F^S_{rv} \tag{3.12a}$$

$$\text{s.t.} \sum_{r \in R^w_v} \beta_{rv} \leq 1 \qquad \forall v \in V \tag{3.12b}$$

$$\sum_{v \in V_r} \beta_{rv} = 1 \qquad \forall r \in R^w \tag{3.12c}$$

$$\beta_{rv} \in \{0, 1\} \qquad \forall r \in R^w \;\; \forall v \in V_r \tag{3.12d}$$

This integer linear optimization problem is unimodular. Hence, the linear problem resulting from a relaxation of the integrality condition (3.12d) ($0 \leq \beta_{rv} \leq 1 \; \forall r \in R^w \;\; \forall v \in V_r$) will return the same results and solution approaches for typical linear problems can be applied.

The optimization problem considers only the re-assignment of requests that are waiting to be picked up. The solution can have several consequences on vehicle movements.

- A vehicle that was idle and is assigned to a request r starts driving towards x^p_r immediately.
- A vehicle that is en-route to pick-up a request r' and re-assigned to r is instantly redirected towards x^p_r.

- A vehicle, which is en-route to drop off request r'' and was assigned to pick up request r' after, still remains en-route to drop off r'' if r is assigned to it instead of r'.
- When a vehicle v, which had an assignment prior to the re-optimization, is without assignment after the re-optimization, it becomes idle and waits for further instructions.

The frequency of assignment re-optimizations has to be aligned with the computational time required to create V_r for all active requests and solve problem (3.12). For large scale fleets with thousands of vehicles, the time between re-optimizations Δt^R is typically chosen in the range of $10 - 30$ seconds.

To reduce the computational burden, so-called *RV-heuristics* can be applied. These heuristics remove some (r, v) edges in the bipartite matching graph, which are likely not part of the optimal solution. In the hailing case, this is easily implemented by only keeping the $N^{RV,h}$ vehicles with the lowest control objective value for each request r.

3.3.2 Creation of Offers and Re-Optimization of Vehicle Plans (Pooling)

In a pooling service, a request's drop-off stop at x_r^d does not have to follow the pick-up stop x_r^p immediately. Therefore, the set of requests waiting for a vehicle R_v^w can consist of multiple requests and there can be multiple possible sequences of pick-up and drop-off stops. Figure 3.3 illustrates the generation of all possible vehicle plans with a new request r for an existing vehicle plan with one on-board request and an existing vehicle plan with one waiting request.

The number of possible stop combinations grows rapidly. As shown in Table 3.1, even for a single vehicle with $|R_v^o|$ on-board and $|R^w|$ waiting requests, the number of possibilities can go into the thousands for just a few requests.

Therefore, computationally efficient algorithmic approaches to find feasible routes are necessary. A first criterion to eliminate possible vehicle plans is to check if a vehicle v is close enough to serve a request r. The backwards-Dijkstra search presented in section 3.3.1 can be utilized again, but with a different definition of vehicle availability. Because the order of stops can be re-arranged, it is theoretically always possible to insert the pick-up stop x_r^p for a new request r at the first available position in each vehicle's plan:

- Idle vehicles can start to drive to x_r^p right away
- Vehicles, which are currently stopping for boarding, can be directed to x_r^p after finishing the boarding process
- Vehicles, which are en-route to another stop or idle, can be redirected towards x_r^p

In conclusion, the availability (x_v, t_v) of each vehicle is determined by its current location x_v, and either the current time or the time it finishes a current boarding process. If the vehicle cannot pick up request r in time, i.e., $t_v + t[x_v \to x_r^p] > \tau_r + \tau_r^w$, the vehicle does not have to be considered to serve r.

In contrast to the hailing case, this is not a sufficient criterion to find a feasible vehicle plan with the insertion of r. The minimum requirement is that equations (3.7) are satisfied

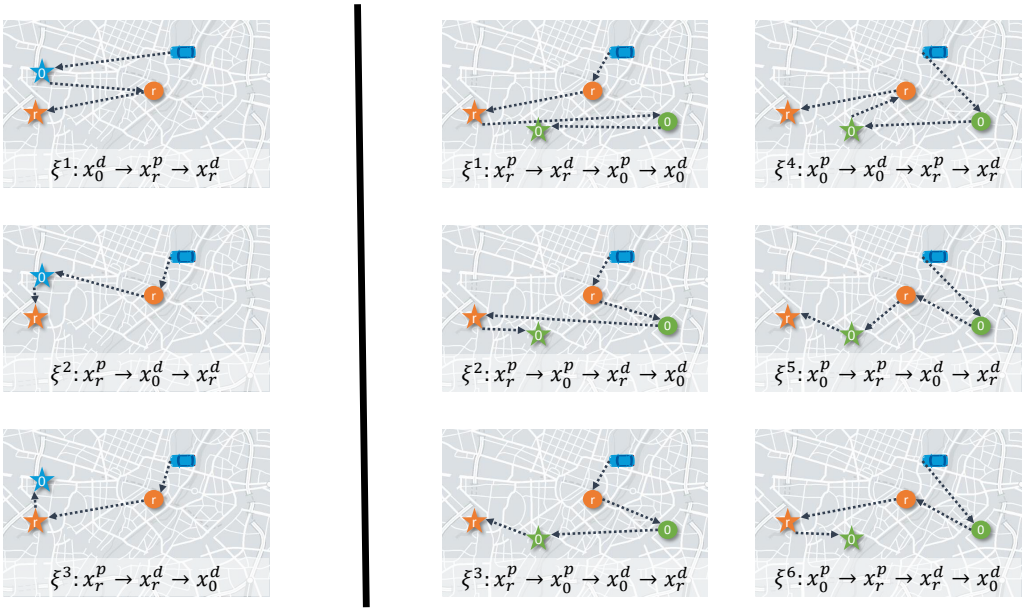


Figure 3.3: Possible insertions of stops for request r into an existing vehicle plan with one on-board request (left) and into an existing vehicle plan with one waiting request (right). Circles and stars illustrate request pick-up and drop-off locations, respectively.

for the on-board requests, i.e $R_v^\xi = R_v^o \cup \{r\}$. The concept and definition of stops X_s is computationally more efficient than creating a vehicle plan and checking all conditions (3.7) for each request and each part of the route after. Instead, the time constraints can be checked stop-by-stop during the creation process. The set of boarding and disembarking requests (R_s^+ and R_s^-) for a given stop s remain the same regardless of the sequence of stops. Hence, the latest arrival time at a stop t_s^l reflects both the waiting time conditions of all requests waiting to board at x_s and all in-vehicle/detour time conditions of requests that want to disembark at x_s. This allows eliminating a lot of sequences: consider a vehicle plan, where the latest arrival time constraint is violated for stop s; then no plan with X_s at a later position in the sequence of stops can be feasible.

The computation of the latest arrival time t_s^l is depending on the stop sequence but can be partially preprocessed for all on-board requests[4]:

$$\bar{t}_s^l = \min\left(\min_{r \in R_v^o}(t_r^p + \tau_r^d), \min_{r \in R_s^+}(\tau_r + \tau_r^w)\right) \tag{3.13}$$

where t_r^p denotes the pick-up time of an on-board request r. The first term considers all detour time constraints of on-board requests and the second considers all waiting time constraints of requests waiting to board the vehicle (R_s^+). As a possible permutation of stop sequences is created (considering that pick-up stops have to be placed before drop-off stops), the expected pick-up times of waiting requests $\bar{t}_r^p$ are computed and the change of the latest arrival time due

[4] A complete preprocessing would be possible if requests would be defined with a latest arrival time instead of a maximum travel or detour time.

$(\lvert R_v^o\rvert, \lvert R^w\rvert)$	(0, 0)	(1, 0)	(0, 1)	(1, 1)	(0, 2)	(2, 0)	(0, 3)	(1, 2)	(2, 1)	(3, 0)	
Possible Routes	0	1	1	3	6	2	90	30	12	6	
$(\lvert R_v^o\rvert, \lvert R^w\rvert)$	(0, 4)	(1, 3)	(2, 2)	(3, 1)	(4, 0)	(0, 5)	(1, 4)	(2, 3)	(3, 2)	(4, 1)	(5, 0)
Possible Routes	2520	630	180	60	24	113,400	22,680	5040	1260	360	120

Table 3.1: Number of possible routes over number of on-board passengers ($|R_v^o|$) and assigned waiting customers ($|R^w|$) for a single vehicle. It is assumed that stops cannot be combined for this computation. The (recursive) code to compute the number of routes can be found in Appendix B.1.

to new in-vehicle time constraints can be adapted dynamically for this specific stop sequence:

$$t_s^l = \min\left(\min_{r \in R_v^w}(\bar{t}_r^p + \tau_r^d), \bar{t}_s^l\right) \tag{3.14}$$

The earliest departure time t_s^e can be derived by the earliest departure time of the boarding requests for each stop regardless of the sequence of stops:

$$t_s^e = \max_{r \in R_s^+}\left(\tau_r^e + t^b\right) \tag{3.15}$$

The feasibility is checked by planing the arrival and departure times at the respective stops. Let t_i be the current time and x_v the location of vehicle v. Then the arrival time t_1^a at the first stop $X_1 = (x_1, t_1^s, t_1^l, t_1^e, R_1^+, R_1^-)$ is determined by the travel time on the fastest route, i.e., $t_1^a = t_i + t[x_v \to x_1]$. If $t_1^a \leq t_1^l$, this stop is feasible. The vehicle can depart from this stop at $t_1^e = \max(t_1^a + t_s, t_1^e)$ if there are other stops. For all later stops s, the arrival time is given by $t_s^a = t_{s-1}^e + t[x_{s-1} \to x_s]$ and so forth.

There can be multiple combinations of distributing requests on vehicles and multiple feasible vehicle plans containing the exact same requests. An operator will select/assign a certain vehicle plan based on a pooling control objective on vehicle plan level. As for the hailing case, the costs for driving and the time required to serve all assigned requests are considered again. Since the complete vehicle plan is no longer given by the sequence of pick-up stops, the complete trajectory information of a vehicle plan ξ_v are necessary to evaluate its control objective function value $F_{\xi_v}^S$:

$$F_{\xi_v}^S = c^D \cdot d_v^\xi + c^{VOT} \sum_{r \in R_v^\xi}\left(t_r^w + t_r^d\right) \tag{3.16}$$

where $t_r^w + t_r^d$ represents the expected waiting and in-vehicle time of all requests R_v^ξ in the vehicle plan ξ_v. t_r^w can also be set to the actual waiting times or be ignored for all on-board requests as they remain constant. This objective function trades off pooling multiple users into a route (if VKT can be saved) and the additional waiting and detour time of the respective users. The choice of coefficients is important. For a too large value c^{VOT}, the AMoD system would predominantly use hailing and only pool users if (i) it means very little time delay for the users (e.g., if they had exactly the same stops) or (ii) if there are no vehicles for exclusive

(n_o, n_w)	(0, 3)	(1, 2)	(2, 1)	(0, 4)	(1, 3)	(2, 2)	(3, 1)	(0, 5)	(1, 4)	(2, 3)	(3, 2)	(4, 1)
IH Routes	15	10	6	28	21	15	10	45	36	28	21	15
% Routes	16.67	33.33	50.0	1.11	3.33	8.33	16.67	0.04	0.16	0.56	1.67	4.17

Table 3.2: Number of routes tested by the insertion heuristic (IH) and its share of all possible routes for a given number of on-board and waiting requests (n_o, n_w).

service left. On the other hand, ignoring c^{VOT} might be disadvantageous for both users and AMoD system. Distances as sole objective can send vehicles into congested areas where even though distance is saved neither the operator nor the users are satisfied.

Creation of Offers (Pooling): As in the hailing case, the backwards-directed Dijkstra search determines the set of vehicles V_r that could satisfy a request's maximum waiting time constraint τ_r^w but with the different definition of vehicle availability. The set V_r is typically larger than in the hailing case (for the same fleet size) as vehicles are available right away whereas in the hailing case current trips to drop-off locations have to be fulfilled first.

As mentioned, the feasibility conditions have to be checked for possible insertions. For simplicity and high computational performance, an insertion heuristic only considers the inclusion of the new requests' pick-up and drop-off stops in the *currently assigned* sequence of stops. This method explores all possible stop combinations for up to two requests and becomes a heuristic from the third request on. Let for example $\psi_v : x_1^p \rightarrow x_2^p \rightarrow x_1^d \rightarrow x_2^d$ be the original route; the addition of r does not consider any vehicle plan that is based on the permutations of these four stops, which considerably reduces the total number of permutations from 90 to 15. The insertion heuristic reduces the maximum number of route combinations significantly; the share of all possible route combinations for up to four previously assigned (on-board or waiting) requests and one new (waiting) request is shown in Table 3.2. The feasibility checks result in a set of feasible vehicle plans Ξ_r for request r, sets of feasible vehicle plans Ξ_r^v for request r and vehicle v, and a set of vehicles with at least one feasible vehicle plan V_r^Ξ.

If $V_r^\Xi = \emptyset$, the operator rejects a request. Otherwise, the control objective function determines, which vehicle and vehicle plan are assigned to a new request. As in the hailing case, the dynamic nature of the problem prohibits a derivation of an optimal choice of this multi-objective function.

Let $\xi_{rv} \in \Xi_r$ be a feasible plan that resulted from the insertion of request r into the currently assigned vehicle plan ξ_v. The local optimization problem of the insertion heuristic finds the vehicle v^* and vehicle plan ξ_v^* that generates the least additional costs when accommodating the stops of the new request r.

$$\xi_v^* = \underset{\xi_{rv} \in \Xi_r}{\arg\min}\, \Delta F_{\xi_{rv}}^S = \underset{\xi_{rv} \in \Xi_r}{\arg\min} \left(F_{\xi_{rv}}^S - F_{\xi_v}^S \right) \tag{3.17}$$

This problem can be formulated as a vehicle assignment problem by defining an intermediary step (without computational effort) that selects the best vehicle plan for each vehicle and then

compares the vehicles:

$$\xi_{rv}^* = \underset{\xi_{rv} \in \Xi_r^v}{\arg\min} \Delta F_{\xi_{rv}}^S \qquad \forall v \in V_r^\Xi \tag{3.18a}$$

$$v^* = \underset{v \in V_r^\Xi}{\arg\min} \Delta F_{\xi_{rv}^*}^S \tag{3.18b}$$

ξ_v^* can then be retrieved as ξ_{rv}^* for vehicle v^*. This approach consisting of (i) finding feasible solutions, (ii) selecting the best solution for a single vehicle, and finally (iii) assigning a vehicle, will be handy when it comes to the re-optimization of pooling assignments.

An important remark is that the realization of the offered waiting and detour time are not guaranteed. In a hailing operator policy, which uses only an insertion heuristic to create offer and make assignments, the offers would remain exact. However, the re-optimization of assignments can change the actually realized pick-up and drop-off times. For the pooling case, realized pick-up and drop-off times could differ from the offer even without global re-optimization. In an operating policy only using the insertion heuristic, a new assignment to a vehicle v can affect all previously assigned requests R_v. Obviously, global re-optimizations can affect the realized pick-up and drop-off times in the pooling case as well. Nevertheless, the described methodology to generate offers is very useful as it allows a quick estimation of level-of-service for users and ensures that requests, which confirmed an offer, will always remain in feasible vehicle plans (for static network conditions). The correctness of the last statement is proven easily as the insertion heuristic only generates feasible solutions and any re-optimization always has the option to not change any vehicle plans.

After receiving the offer, user r makes a decision and several cases can occur. If r declines the offer, the original vehicle plan ξ_v is restored. Otherwise, ξ_v^* will be the new vehicle plan of vehicle v. If the pick-up stop at x_r^p is not on the first position in the sequence of stops ξ_v^*, v continues along its current trajectory. If v was idle, then it starts driving towards x_r^p; if it was en-route to another stop but x_r^p is the new first stop, vehicle v is redirected.

Re-Optimization of Vehicle Plans (Pooling): Periodic optimization-based modification of vehicle plans can help to improve the fleet performance. However, compared to the hailing case, there are more degrees of freedom, which on the one hand offers higher optimization potential, on the other hand also requires much higher computational effort.

For the insertion heuristic, it was possible to define request-vehicle tuples (r, ξ_v) that represent the vehicle plan with the best control function value, which could be assigned to a vehicle. This is no longer possible for the global re-optimization as multiple requests can wait for a vehicle and the sequence of all stops, including the drop-off stops of the on-board requests, can be modified. Hence, the set of active requests R^a contains both the requests waiting for pick-up R^w and the on-board requests R^o.

As indicated in table 3.1, a brute force approach to enumerating every possible combination of plan stops becomes hardly tractable even for a single vehicle and a small number of requests. To address this curse of dimensionality, this thesis follows the approaches developed by Alonso-Mora, Samaranayake, et al. [2017] and Engelhardt, Dandl, and Bogenberger [2019] with some modifications. The concept is to eliminate as many possible stop combinations outright, i.e., without even having to explicitly build the route and check

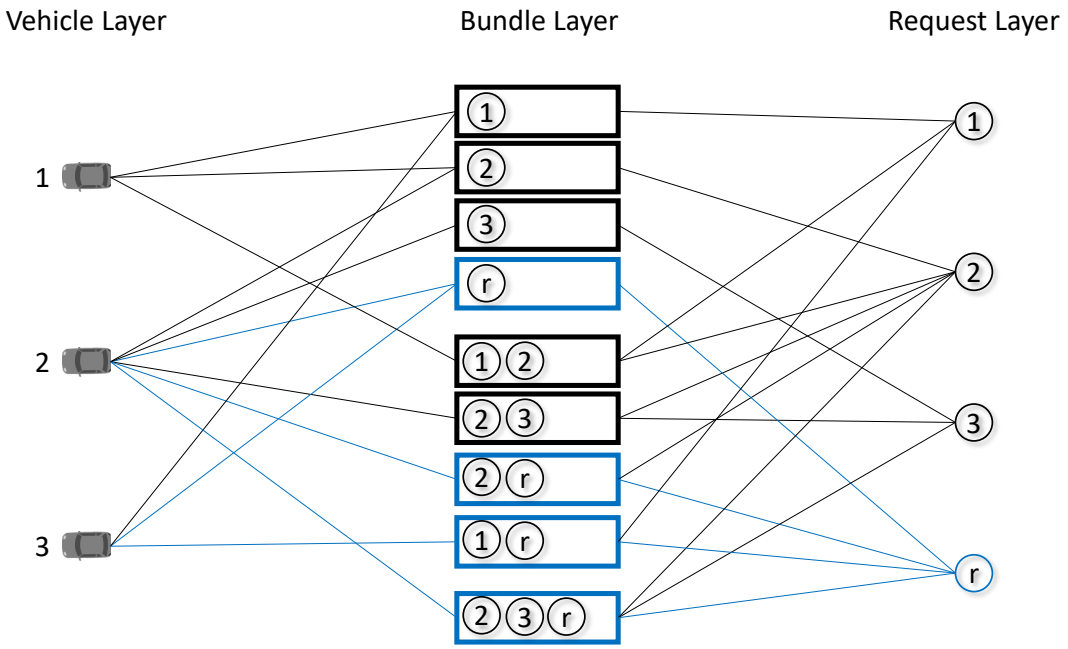


Figure 3.4: Exemplary V2RB-graph. The edges highlighted in blue show possible paths/assignment options for an assignment of request r.

for feasibility, before finding the best combination according to the control objective function F^S.

On a very high level, the idea is to build an extra layer between requests and vehicles, denoted by bundle layer. This layer contains request bundles $b \in B$ which are sets of requests that can be served together based on their time constraints. The tasks are to find the vehicles V_b efficiently that can serve a bundle and then solve a two-layered assignment problem that selects which vehicles should serve which bundles. An important realization at this point has to be that on-board requests do not have to be assigned to a vehicle anymore. Hence, all on-board requests R_v^o are associated with the vehicle v in this graph. Only the requests R^w have to be assigned to a vehicle and are represented by nodes in the request layer.

A graph for the resulting assignment problem is denoted as vehicle to request bundle (V2RB)-graph[5]. Figure 3.4 displays an example. Such V2RB-graph actually represents the complete ride-pooling re-optimization problem and therefore graph elements absorb a lot of complexity.

An edge between a vehicle v and a request bundle b is denoted as V2RB Ξ_{vb}. It contains all feasible vehicle plans of vehicle v that could be used to serve all requests within the bundle b. In the example with 3 vehicles, 4 requests, only the 9 of the $\binom{4}{1} + \binom{4}{2} + \binom{4}{3} + \binom{4}{4} = 15$ possible request bundles are drawn, which produce feasible V2RBs. Moreover, the cost of a V2RB edge is determined by its representative vehicle plan ξ_{vb}, which is the solution of a single vehicle DaRP:

$$\xi_{vb} = \underset{\xi \in \Xi_{vb}}{\arg\min}\, F_\xi^S \tag{3.19}$$

[5] Alonso-Mora, Samaranayake, et al. [2017] use a very similar construct denoted by request-trip-vehicle graph, which is conceptually the same but with slightly different graph elements.

Hence, the definition of a V2RB already contains finding all feasible vehicle plans and the selection of the best single-vehicle plan. Therefore, the edge between vehicle 2 and the request bundle $b = \{2, 3, r\}$ in Figure 3.4 could represent up to 5400 possible vehicle plans if vehicle 2 has two on-board requests, and its costs represents the best control objective function value $F^S_{vb} = F^S_{\xi_{vb}}$ of all of these plans with respect to equation (3.16). It should be noted that all on-board requests have to be considered throughout, i.e., all sequences with the drop-off stops of on-board requests have to be accounted for when searching feasible vehicle plans and also computing the control objective function values. The connections from a request to the bundle layer show all possible bundles B_r that a request r could be served it with. The set B_r always contains the bundle $b = \{r\}$ but can also contain bundles with other requests. Since r was already confirmed, one of the bundles will have to be assigned for each request r.[6]

A key definition is the rank of a bundle and connected V2RBs as the number of requests within a bundle. The graph representation becomes powerful because of one main observation: a V2RB Ξ_{vb} of rank n with request bundle $b = \{r_1, ..., r_n\}$ cannot be feasible if any of the n V2RBs of rank $n - 1$, where one of these requests is not included, is not feasible. Moreover, only insertions of the stops of request r_n into feasible vehicle plans of the V2RB $\Xi_{vb'}$ with request bundle $b' = \{r_1, ..., r_{n-1}\}$ can be part of Ξ_{vb}. Hence, a meaningful approach to finding all feasible V2RBs increases in rank and the graph-building process can be run in parallel for each vehicle v.

Since the addition of more and more boarding processes and detours decrease the probability of finding feasible V2RBs and vehicle plans for an AMoD service with high user convenience, the graph-building process can become tractable for fleets with hundreds to thousands of vehicles [Alonso-Mora, Samaranayake, et al., 2017; Engelhardt, Dandl, and Bogenberger, 2019], especially if additional measures are introduced for the rank 1 and 2 V2RBs.

For rank 1 V2RBs, i.e., single waiting requests, the backwards-Dijkstra method can be utilized to filter vehicles that cannot reach a request r before its waiting time expires. After that, it has to be checked whether r's stops can be inserted without violating the on-board requests. The remaining edges between vehicles and single-request bundles are denoted RV-graph. This graph builds the basis for the two sets B_v for all bundles that are connected to a vehicle and V_b for all vehicles that are connected to a bundle.

For the step to rank 2 V2RBs, it is beneficial to mark requests that cannot be pooled due to their time constraints. This can be very beneficial for requests with similar pick-up locations, as they might have many RV graph connections to the same vehicle. However, as their time constraints are incompatible, the combination of these two requests would be checked and found to be not feasible for each single vehicle. To avoid this, an RR-graph is introduced, which connects two requests r_1 and r_2 if any of the 6 possible routes (with only these two waiting requests) of a hypothetical vehicle starting from x^p_1 or x^p_2 would be feasible. If this is not the case, no real vehicle will be able to serve the bundle $b = \{r1, r2\}$.

If the problem instances are still too large, heuristics could be utilized to, for example,

[6] Alternatively, a large assignment reward could be added for each request by putting a large negative cost on all edges between bundles and requests. The edges to different requests in the V2RB-graph could also have different weights to allow the prioritization of certain requests, e.g. confirmed over non-confirmed requests if both should be treated in the procedure.

limit the possible insertions in the single-vehicle DaRP problems [Alonso-Mora, Samaranayake, et al., 2017; Simonetto et al., 2019] or remove edges from the RV-graph randomly [Simonetto et al., 2019] or based on criteria [Engelhardt, Dandl, and Bogenberger, 2019]. These heuristics and the investigation of computational gain vs loss of optimality are not within the scope of this thesis.

Once the V2RB-graph is built, the following assignment problem can be formulated and solved:

$$\min_{\boldsymbol{\beta}} \sum_{v \in V} \sum_{b \in B_v} \beta_{vb} \cdot F^S_{vb} \tag{3.20a}$$

$$\text{s.t.} \sum_{b \in B_v} \beta_{vb} \leq 1 \qquad \forall v \in V \tag{3.20b}$$

$$\sum_{b \in B_r} \sum_{v \in V_b} \beta_{vb} = 1 \qquad \forall r \in R^w \tag{3.20c}$$

$$\beta_{vb} \in \{0,1\} \qquad \forall v \in V \;\; b \in B_v \tag{3.20d}$$

The binary variables $\boldsymbol{\beta}$ indicate the selected V2RB assignments. If $\beta_{vb} = 1$, the representative vehicle plan ξ_{vb} is the new vehicle plan of vehicle v. As in the hailing case, equation (3.20b) limits the number of assigned vehicle plans to at most one. Equation (3.20c) ensures that each request is part of an assigned vehicle plan; visually speaking, exactly one of the paths starting from each request node (e.g. the ones highlighted in blue for request r in Figure 3.4) has to be assigned. A changed vehicle plan can cause a vehicle to stop (previously existing vehicle plan was deleted), be redirected (vehicle was en-route and has a new first stop), or continue as planned (first stop remains the same).

The computational effort to create the complete V2RB-graph and solve the subsequent optimization problem is very high. To create solutions in limited time in the range of 10-30 seconds, techniques like parallelization and RV-heuristics are useful and necessary for large-scale problems. The vehicle-search processes can be parallelized for each request and the V2RB-graphs can be built for each vehicle in parallel. RV heuristics in the V2RB-graph building process can be applied in multiple steps. Since the V2RB-graph building process is often computational at least as expensive as solving the optimization problem [Engelhardt, Dandl, and Bogenberger, 2019], removing edges from the final V2RB-graph brings only minor computational benefits. Nevertheless, finding the exact solution of the assignment problem of equation (3.20) might last a long time and the introduction of an optimization time out can be useful[7]. Another option would be to reduce the amount of vehicle plans within a V2RB (e.g. by removing the vehicle plans with the least utility or a heuristic insertion like Tabu search in the first place). Finally, the most computation time can be gained by removing vehicles from V_r before all insertions and feasibility checks are performed for a limited set of vehicles.

This thesis applies two complementing RV heuristics of the last sort. The first RV heuristic aims to guarantee service if possible. Therefore, the vehicles V_r are scanned for their workload and the $N^{RV,wl}$ vehicles with the least workload measured in number of currently assigned

[7] Many commercial solvers, such as CPLEX or Gurobi offer such an option.

stops. Therefore, this heuristic favors the use of idle vehicles which provides a high probability to find a vehicle satisfying the time constraints of new requests. On the downside, these vehicles are not that likely to find solutions with higher degrees of pooling. In contrast, the second RV heuristic aims to find vehicles that present a high probability of pooling. For this purpose, the alignment of the directions defined by the currently assigned vehicle plan and the new request's trip are measured by

$$align_{rv} = \vec{e_v} \circ \vec{e_r} \tag{3.21}$$

where $\vec{e_v}$ is the unit vector from the current vehicle location to the location of the last currently assigned stop and $\vec{e_r}$ is the unit vector from r's origin to destination. The heuristic then selects the $N^{RV,al}$ vehicles in V_r with the highest $align_{rv}$ values.

Moreover, it should be mentioned that the application of RV heuristics are also meaningful to decrease response time when creating pooling offers. Hence, these RV heuristic limiting the number of vehicles, for which all possible insertions are explicitly checked for feasibility, are applied in the offer phase and the re-optimization phase. However, it is important that additional care is taken when applying heuristics in the re-optimization process. For each request, the currently assigned vehicle is added to the list of vehicles, for which the V2RB is created, in order to guarantee that the solution after re-optimization is at least as good as the solution before the optimization.

3.3.3 Handling of Reservations

Reservations cannot be treated the same way as immediate-service requests. In the developed hailing re-optimization strategy, the assumption that a vehicle plan should only contain one pick-up does not hold anymore if another pick-up of a reservation is half an hour or longer in the future. Therefore, the V2RB framework is used for both hailing and pooling; however, vehicle capacity is limited to $\rho_v = 1$ in the hailing case.

In the V2RB framework, a lot of vehicles will be able to reach a new reservation request and a lot of routing possibilities remain available making the developed pooling re-optimization methodology intractable for a relatively small number of requests (compared to its online capability). Depending on the urgency of the response, reservations can be classified as short-term or mid-term decisions. On the one hand, they require more foresight, on the other hand they should be treated as short-term decision if an immediate accept/reject response is required as is the case in this thesis.

Therefore, another strategy is developed to treat a reservation request r with $\tau_r^e > \tau_r$. The idea is displayed in Figure 3.5. After an initial insertion heuristic checks the feasibility and makes an initial assignment of request r to vehicle v_r, the request is kept as a constraint in the possible routes of vehicle v_r in future re-optimization time steps. For pooling re-optimizations, the request is essentially treated like an on-board request: r is not part of any request bundle but the stops related to the pick-up and and drop-off of r have to be considered in any vehicle plan of v_r when the feasibility of V2RBs is checked. This essentially means for a hailing operation that the assignment of an immediate request r' is only possible if the vehicle can serve r' and reach x_r^o before $\tau_r^e + \tau_r^w$ passes.

r remains as a constraint for V2RB until it becomes a variable of the global re-optimization at time t_i, when the earliest pick-up time is within the short-term horizon, i.e., $t_i + T^S \leq \tau_r^e$.

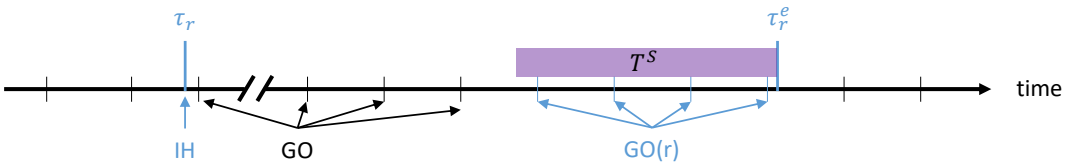


Figure 3.5: Treatment of a reservation request r. An initial insertion heuristic (IH) assigns r to a vehicle. The subsequent global re-optimizations (GO) keep this request r in a feasible route of the assigned vehicle. When the earliest pick-up time τ_r^e is within the short-term horizon of a re-optimization step GO(r), r is considered in this re-optimization.

In the global re-optimizations denoted by GO(r) in Figure 3.5, r is treated as any immediate request, which can be re-assigned to other vehicles.

The control objective function (for both insertion heuristic and global re-optimization) is slightly modified to account for reservation requests. The driving cost term $(c^D \cdot d_v^\xi)$ and the value-of-time term $(c^{VOT} \cdot (t_r^w + t_r^d))$ still are meaningful. However, it also makes sense to consider the time a vehicle would have to wait for the reservation request r at x_r^p before departing. In many cases, it might make sense to assign a vehicle with low downtime instead of one with a long waiting time. For instance, when the reservation request r wants to be picked up in half an hour, a vehicle finishing at x_r^p in 29 minutes is likely to remain the vehicle to serve r. In contrast, a vehicle finishing at x_r^p in 15 minutes would have enough downtime to fulfill different tasks but might be constrained by its task to return to x_r^p for the reservation request.

Let t_v^w denote the time a vehicle would have to wait for customers at stops s in a certain vehicle plan ξ_v. t_v^w can be computed for a single stop by the difference between the earliest departure time t_s^e and the expected vehicle arrival time t_s^a (considering the service duration t_s^s it anyway should stop there for boarding $t_s^s = t^b$):

$$t_v^w = \sum_{X_s \in \xi_v} (t_s^e - (t_s^a + t_s^s)) \tag{3.22}$$

Note that $t_s^e - (t_s^a + t_s^s) = 0$ for pick-ups related to immediate service requests. However, for reservation requests, the earliest pick-up time constraint can lead to $t_v^w > 0$. Hence, the control objectives F^S can be modified to account for the time vehicles have to wait for reservation requests[8]:

$$F^S(\text{with reservations}) \leftarrow F^S(\text{without reservations}) + c^{VOT} \cdot t_v^w \tag{3.23}$$

The vehicle search process, which was an essential step to cut the solution space for online requests, can be modified and utilized to make reasonable cuts for reservation requests as well. Instead of using the current availability as described in the previous subsection, the availability (i.e., position and time of availability) of each vehicle is determined by the last stop, which is planned to take place before τ_r^e of a new reservation request r. The vehicle-search procedure

[8]

(backwards-Dijkstra) can then be limited in space to select vehicles that anyway make a stop near the new reservation request.

The advantages of the described approach are:

- the feasibility of a reservation can be checked immediately and the operator can instantly accept/reject a reservation request
- a reservation confirmed by the operator is guaranteed to remain feasible
- the vehicle assignment becomes part of the optimization problem approximately at the time when a vehicle actually has to drive for a pick-up
- the computational effort of reservations is similar to those of online requests

This approach cuts the solution space severely and a fully static DaRP solution could provide better solutions but at a significantly higher if not infeasible computational cost.

3.4 Tactical Mid-Term Decisions

In the last section, algorithmic approaches were developed (i) to respond to app requests asking for immediate or future service with realistic offers and (ii) to re-optimize current vehicle plans globally. These decisions were based on route forecasts in the range of the maximum waiting time and trip time of requests. This section explains the utilization of demand and supply forecasts to improve fleet performance by planning further ahead of time even without the knowledge of explicit reservation requests.

3.4.1 Demand and Supply Forecasting

Stochastic knowledge of future demand represents a new set of state variables $\Lambda(t_i)$, which are not handled by the short-term decisions in the framework of this thesis. These forecasts are generally available in a spatially and temporally aggregated form. They typically refer to certain time horizons T and a set of zones Z; they can be predictions of origin-destination trips (Λ_{od}), just trip departures (Λ_z), or just trip arrivals ($\bar{\Lambda}_z$), where $od \in Z^2$ and $z \in Z$. The forecasts are typically based on historic data and can be improved by online methods [Sayarshad and J. Y. J. Chow, 2016]. Nowadays, machine-learning techniques are applied to generate high-accuracy demand predictions[9].

While there are established methods for demand predictions, forecasting the AMoD supply side is much more complex. In the short-term the current state with the current positions and all its currently assigned vehicle plans represents a good estimator; however, the further into the future the prediction goes, the less valid this approach becomes. Far in the future, most vehicles will have received new assignments and vehicle availability will be determined by demand forecasts. In the hailing case, vehicles will become available at the departure locations; hence, forecast of trip arrivals per zone $\bar{\Lambda}_z$ becomes a good estimator for the number of vehicles

[9] Didi claims that their 15-minute predictions have 85% accuracy: `https://www.didiglobal.com/news/newsDetail?id=323&type=news`. However, the used spatio-temporal resolution is not clear.

that will become available in zone z in the respective time horizon. The author is unaware of a methodology that describes and evaluates the transition from the current state with existing plans to future supply forecasts. A consistent methodology that does not use the current assignments at all is to look at the stack of idle vehicles and use departure and arrival forecasts to update it (see e.g. [Carron et al., 2019; Charkhgard et al., 2020] but this method does not consider the exact knowledge of current vehicle plans in the near future. Hence, such an approach is consistent with an assignment of idle vehicles only, which however performs much worse than AMoD systems considering current vehicle plans [M. Hyland and Mahmassani, 2018a; Dandl and Bogenberger, 2018].

Another unanswered question is how forecasts of vehicle demand should be treated in the pooling case; the amount of required vehicles will be lower than the number of trip departures because of pooling. Inter-zonal relationships with high pooling probabilities could be considered. However, these pooling adaptations are out of the scope of this thesis. Instead, the same methodology is used for hailing and pooling. This implicitly assumes that both the number of arriving vehicles and the number of required vehicles will scale similarly with the pooling rates. The method for vehicle imbalance forecasts without any operator actions is described in this section and will be extended by repositioning and dynamic pricing in the subsequent sections.

The developed strategy divides the mid-term time horizon T^M into multiple equidistant steps. Let $T^{M,1}$ be the first forecast step, i.e., a time horizon starting from t_i with duration $\Delta T^M \approx \mathbb{E}[t^d]$ approximately representing the average trip duration. The strategy assumes that the forecast of vehicle availability based on the current vehicle plans is valid until $T^{M,1}$. Therefore, the imbalance I_z^1, which denotes the surplus or deficit of vehicles within $T^{M,1}$, is given by

$$I_z^1 = |V_z^I| + |V_z^{P,1}| - \Lambda_z^1 \tag{3.24}$$

where $|V_z^I|$, $|V_z^{P,1}|$ and Λ_z^1 refer to the number of currently idle vehicles in zone z, the number of vehicles that end their current vehicle plan within zone z and the estimated trip departures from zone z in the time horizon $T^{M,1}$, respectively. If demand forecasts are given in OD form, the predicted trip departures in zone z and forecast step j can simply be calculated from

$$\Lambda_z^j = \sum_{d \in Z} \Lambda_{zd}^j \tag{3.25}$$

The remainder of the mid-term horizon T^M is divided into several steps of the same size ($T^{M,2}$, $T^{M,3}$, etc.) as illustrated in Figure 3.6. The imbalance in zone z within time horizon $T^{M,j}$ is given by:

$$I_z^j = \max(I_z^{j-1}, 0) + \bar{\Lambda}_z^j - \Lambda_z^j \quad \forall j > 1 \tag{3.26}$$

The first term in equation (3.26) (represented by grey vehicles in Figure 3.6) states that an earlier vehicle surplus can be utilized to absorb a surge in demand. $\max(I_z^{j-1}, 0)$ describes the stack of remaining (idle) vehicles from the previous forecast step $j-1$. In contrast, unsatisfied demand, i.e., users that likely have not been served due to a vehicle shortage and already left the system, does not have to be considered in future forecast steps.

If the forecasts are given in OD form, the trip arrival estimates can be calculated from

$$\bar{\Lambda}_z^j = \sum_{j' \leq j} \sum_{o \in Z} \Lambda_{oz}^{j'} \delta_{oz}^{j',j} \tag{3.27}$$

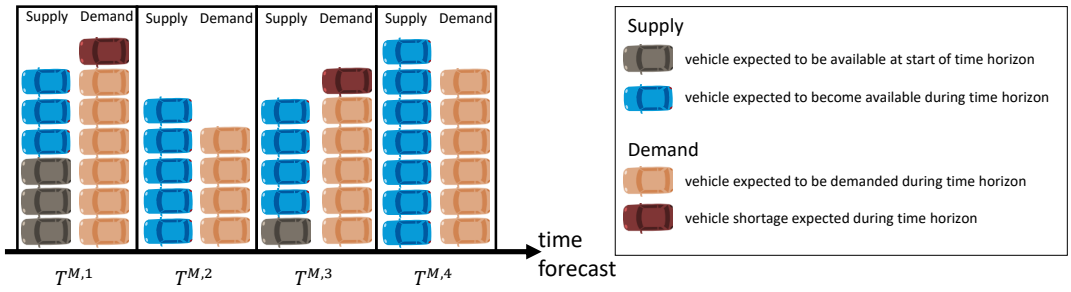


Figure 3.6: On forecasts of vehicle imbalance in this thesis. The blue vehicles represent the expected incoming vehicles based on the current vehicle plan for $T^{M,1}$ and the expected incoming vehicles from earlier trips for all later time horizons.

where $\delta_{oz}^{j',j}$ filters trips starting in time horizon $T^{M,j'}$ that are expected to end in zone z during the time horizon $T^{M,j}$. Thereby $\delta_{oz}^{j',j}$ models the duration that vehicles are not available due to traveling from zone o to zone z. This tensor is derived from network travel times between the centroids of zone o and zone z:

$$\delta_{oz}^{j',j} = \begin{cases} 1 & \text{if } (j-j') \cdot \Delta T^M \leq t[x_o \to x_d] \leq (j-j'+1) \cdot \Delta T^M \\ 0 & \text{else} \end{cases} \tag{3.28}$$

Figure 3.6 and equation (3.26) describe a methodology that treats each zone independently. While this approach is justified for station-based systems, it makes sense to consider the proximity of zones in the determination of imbalances for AMoD systems. If the forecast zones are smaller than the radius a vehicle can reach within the typical user waiting times, vehicles can serve requests from nearby zones thereby potentially making repositioning or dynamic pricing actions not necessary. To address this issue, Syed, Dandl, Kaltenhäuser, et al. [2021] developed the concept of reachability by utilizing kernels to distribute the imbalance between neighboring zones. Compared to this paper, this thesis uses a slightly different density definition and extends the repositioning framework to multiple temporal steps.

The reachability concept assumes that the probability of a request r being assigned to a specific vehicle v is highest at its location of availability, decreases with the distance to this location, and reaches 0 when the distance represents the maximum waiting time, which is denoted by bandwidth h. For simplicity, this is modeled by a linearly decreasing kernel

$$K(x, x_v) = \frac{3}{\pi h^2} \max\left(\frac{1-|x-x_v|}{h}, 0\right) \tag{3.29}$$

where x represents the point of interest for which the imbalance should be computed, x_v is the location of the vehicle, and the factors $\frac{3}{\pi h^2}$ are chosen such that

$$\int_\Omega K(x, x_v)\, d\Omega = 1 \tag{3.30}$$

Assuming a symmetry between demand and supply, this linear kernel density definition is also used to represent the probability of a vehicle v being assigned to a specific request r:

$$K(x, x_r^p) = \frac{3}{\pi h^2} \max\left(\frac{1-|x-x_r^p|}{h}, 0\right) \tag{3.31}$$

Hence, the (supply and demand) imbalance density during time horizon T considering reachability can be written as

$$K^I(x, \boldsymbol{x_v}, \boldsymbol{x_r^p}, j) = \sum_{v \in V^j} K(x, x_v) - \sum_{r \in R^j} K(x, x_r^p) \tag{3.32}$$

where V^j and R^j denote the set of available vehicles and revealed requests expected during forecast step j.

With the assumption that demand and supply are uniformly distributed within the zones[10], the spatio-temporal imbalance density $K^I(x, j)$ can be computed from the expected available vehicles and trip requests of all zones Z, which allows a connection with the previously defined zone imbalance I_z^j.

$$K^I(x, j) = \sum_{z \in Z} \int_{\Omega_z} K^I(x, \boldsymbol{x_z}, \boldsymbol{x_z}, j)\ dx_z \tag{3.33a}$$

$$= \sum_{z \in Z} \left(\sum_{v \in V_z^j} 1 - \sum_{r \in R_z^j} 1 \right) \frac{1}{|\Omega_z|} \int_{\Omega_z} K(x, x_z, j)\ dx_z \tag{3.33b}$$

$$= \sum_{z \in Z} I_z^j \cdot K_z^I(x) \tag{3.33c}$$

Here, the difference of their cardinalities is defined as the expected zone imbalance:

$$I_z^j = \left(\sum_{v \in V_z^j} 1 - \sum_{r \in R_z^j} 1 \right) = |V_z^j| - |R_z^j| \tag{3.34}$$

It is important to note that the density kernel for each zone $K_z^I(x)$ is independent of the imbalance value and can be preprocessed, which makes this method computationally very efficient.

This continuous imbalance density function could be used for machine-learning based approaches. This thesis will apply zone-based optimization algorithms for repositioning. Therefore, the continuous density can be transformed to a corrected zone imbalance $\bar{I}$, which considers reachability, and can be computed from the original zone imbalance values:

$$\bar{I}_z^j = \int_{\Omega_z} K^I(x, j)\ dx \tag{3.35a}$$

$$= \int_{\Omega_z} \sum_{z' \in Z} K_{z'}^I(x) \cdot I_{z'}^j\ dx \tag{3.35b}$$

$$= \sum_{z' \in Z} I_{z'}^j \cdot \left[\int_{\Omega_z} K_{z'}^I(x)\ dx \right] \tag{3.35c}$$

$$= \sum_{z' \in Z} I_{z'}^j \cdot K_{z,z'} \tag{3.35d}$$

[10] This is different from Syed, Dandl, Kaltenhäuser, et al. [2021], where all demand and supply were assumed in the zone centroid.

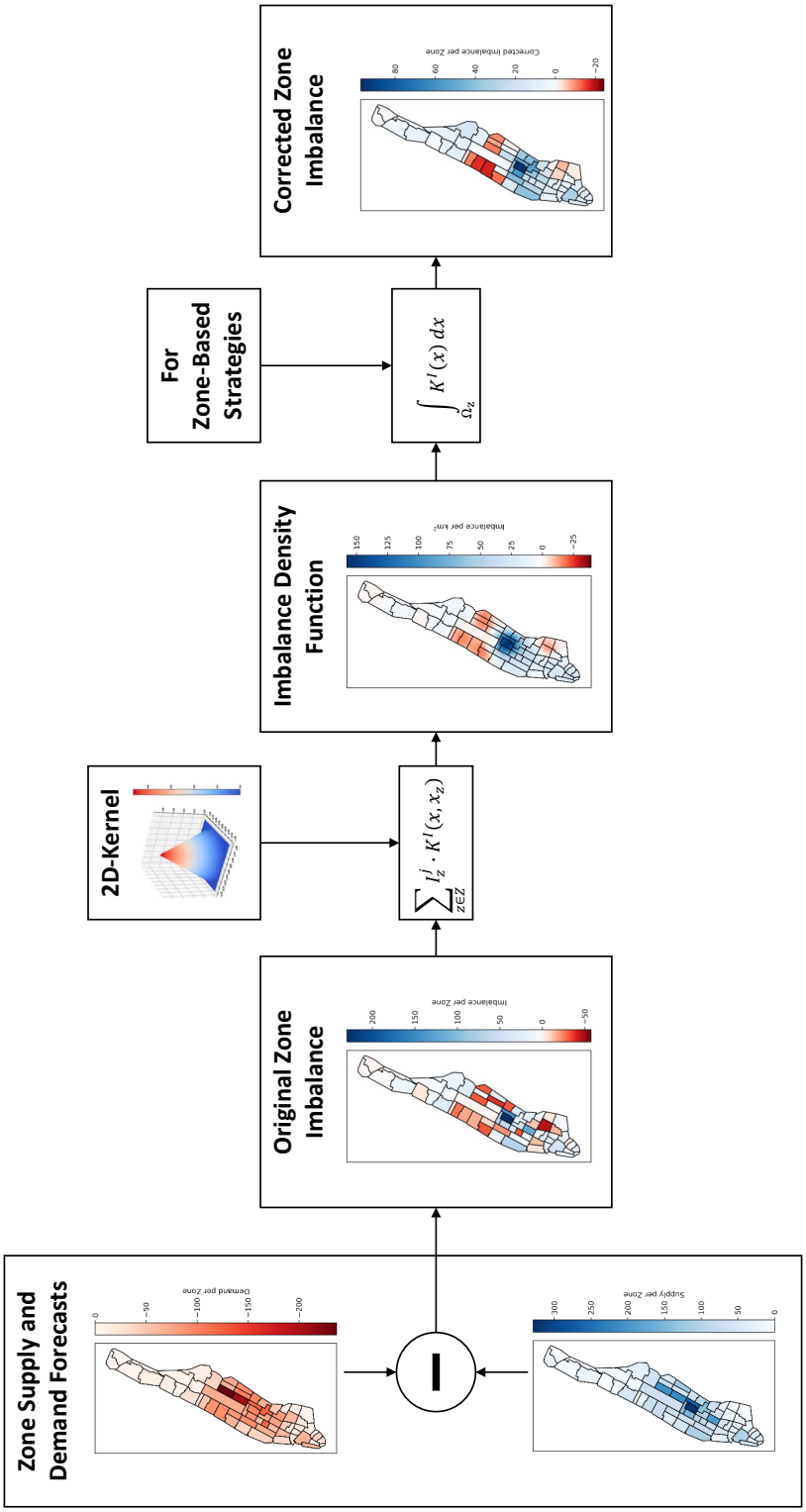


Figure 3.7: On the computation of the imbalance density function. A part of the figure is adapted from Syed, Dandl, Kaltenhäuser, et al. [2021].

These equations highlight the benefit of the uniformity assumption: the reachability correlations $K_{z,z'}$ can be completely preprocessed between all zone-pairs (z, z').

Figure 3.7 illustrates the steps to compute the imbalance density function and the corrected zone imbalance. The correlations between zones are the larger the smaller the zones. Similarly, a larger bandwidth will also increase the correlations. On the other end, the correlation matrix is the unit matrix ($K_{z,z'} = \delta^{z,z'}$) for zero bandwidth. It can be observed in Figure 3.7 that the method distributes imbalances of zones to its neighbors. Therefore, the corrected zone imbalance definition considers the spatial dimension of imbalances. As a consequence, repositioning trips, which would likely be suggested between these nearby zones based on the original zone imbalance, can be avoided.

In the following, repositioning and dynamic pricing strategies are developed on the base of these forecast methods. The goals of these strategies are to (i) try to reposition vehicles in order to avoid vehicle shortages in all steps and (ii) use dynamic pricing to match the price-sensitive demand to the rather well known vehicle supply in $T^{M,1}$. This strategy is suggested as vehicles starting to reposition at the time of decision-making do not necessarily become available during the first time horizon $T^{M,1}$. Hence, the idea is to provide service when possible and use pricing when necessary.

3.4.2 Repositioning

In this section, new repositioning methods are developed. The novelty is the combination of two key ideas: the density-based approach from Syed, Dandl, Kaltenhäuser, et al. [2021] and a multi-time-step model-predictive control approach similar to Charkhgard et al. [2020].

For a multi-step approach, the time horizon T^M of mid-term forecasts is divided into a number of equidistant steps N^{T^M} (represented by bins $T^{M,j}$ with $j \in \{1, ...N^{T^M}\}$ in Figure 3.6) of duration $T^{M,1}$. Moreover, the frequency of making repositioning decisions Δt^M has to be chosen appropriately. For reasons of consistency, it makes sense to choose $\Delta t^M = T^{M,1}$. Thereby, repositioning decisions are valid for the current step and the plans for all future steps can be revisited in the next decision-process. There is a trade-off for the choice of Δt^M: On the one hand, larger values stabilize forecasts. On the other hand, small values allow fast reactions to unexpected changes in demand or supply but likely increase the errors of forecasts and require more expensive computations as more steps are necessary to cover the same time horizon T^M.

In this thesis, the forecast step size $T^{M,1}$ is chosen to be in the range of the average trip duration $\mathbb{E}[t^d]$ as this seems a natural scale to progress in time. The number of forecast steps should be chosen large enough for repositioning trips to cover the whole operating area because the effects of repositioning trips from one end to the other should be considered. Hence, N^{T^M} is defined by the ratio of longest trip duration by average trip duration.

$$N^{T^M} = \left\lceil \frac{\max_{od \in Z^2} t_{od}}{\mathbb{E}[t^d]} \right\rceil \tag{3.36}$$

where $t_{od} = t[x_o \rightarrow x_d]$ denote the fastest route between two centroids. As the number of forecast steps has to be an integer and forecasts for the impacts of the longest possible repositioning trips should be considered as well, the value is rounded up, symbolized by $\lceil . \rceil$.

In each of the forecast steps $j \in J = \{1, ...N^{T^M}\}$, the operator could decide to reposition vehicles between zones. Let $\beta_{od}^j \in \mathbb{N}^+$ be an integer variable denoting the number of vehicles that should be repositioned from zone $o \in Z$ to zone $d \in Z$ at forecast step j. This thesis applies a repositioning strategy, where previously scheduled repositioning trips and the corresponding stops have the $lock$ flag and cannot be changed. Therefore, their effects are already considered in the original forecasts $I_z^j|_{\boldsymbol{\beta}=0}$ before new repositioning decisions are made. Moreover, this implicates that only idle vehicles are considered for repositioning.

The repositioning of vehicles related to β_{od}^j will change the imbalance in zones o and d at the expected arrival time of the vehicle, which is possibly in another forecast step. The departure of a vehicle instantly reduces the number of available vehicles in zone o but the arrival at zone d might be delayed ($j' > j$), which is represented by $\delta_{oz}^{j',j}$ (as defined in equation (3.28)):

$$I_z^j = I_z^j(\boldsymbol{\beta}) = I_z^j\big|_{\boldsymbol{\beta}=0} - \sum_{d \in Z} \beta_{zd}^j + \sum_{o \in Z} \sum_{j'=1}^{j} \beta_{oz}^{j'} \delta_{oz}^{j',j} \tag{3.37}$$

where $I_z^j|_{\boldsymbol{\beta}=0}$ is given by equations (3.24) and (3.26).

The objectives of repositioning are (i) to bring idle vehicles or vehicles that are expected to be idle in the future to zones with negative imbalance, i.e., avoid vehicle shortages ψ_z^j for all zones $z \in Z$ and forecast steps $j \in J$, and (ii) to do so with the least amount of VKT possible. Fig. 3.8 illustrates that bringing additional vehicles to a zone z can decrease a shortage in forecast step j or further in the future ($j' > j$). Hence, the operator predicts to generate additional revenue by reducing opportunity costs. These opportunity costs for not equalizing the vehicle deficit (considering reachability) at j in zone z are given by

$$\gamma^j \bar{f} \cdot \psi_z^j = \gamma^j \, \mathbb{E}\left[f_r\right] \cdot \max\left(0, -\bar{I}_z^j\right) \tag{3.38}$$

$\bar{f} = \mathbb{E}[f^B + (f^D - c^D) \cdot d_r]$ denotes the expected profit, i.e., revenue minus driving costs, for serving an additional request. In practice, this term could also be specific to zone z and forecast time j based on historic trip data originating from that zone during that time horizon. As in the original Bellman equation, γ is a discount factor that exponentially dampens the impact of future rewards. Equation (3.38) assumes that creating additional vehicle surplus does not create additional revenue. Each repositioning vehicle as consequence of $\beta_{od}^j > 0$ also generates costs. The explicit distance costs d_{od} related to a vehicle repositioning from zone o to zone d are approximated by the fastest route distance between the two zone centroids.

Therefore, repositioning is a multi-objective problem, which aims to reduce vehicle deficits with little vehicle movements. In a multi-step formulation, the objectives read:

$$\min \sum_{j \in J} \sum_{z \in Z} \bar{f} \cdot \gamma^j \cdot \psi_z^j \quad , \quad \min \sum_{j \in J} \sum_{o \in Z} \sum_{d \in Z} c^D \cdot d_{od} \cdot \beta_{od}^j \tag{3.39a}$$

The two objectives of this mixed integer problem are in general not aligned as the solution of the distance objective is to not reposition at all, which will usually not be the solution of the vehicle-deficit minimization.

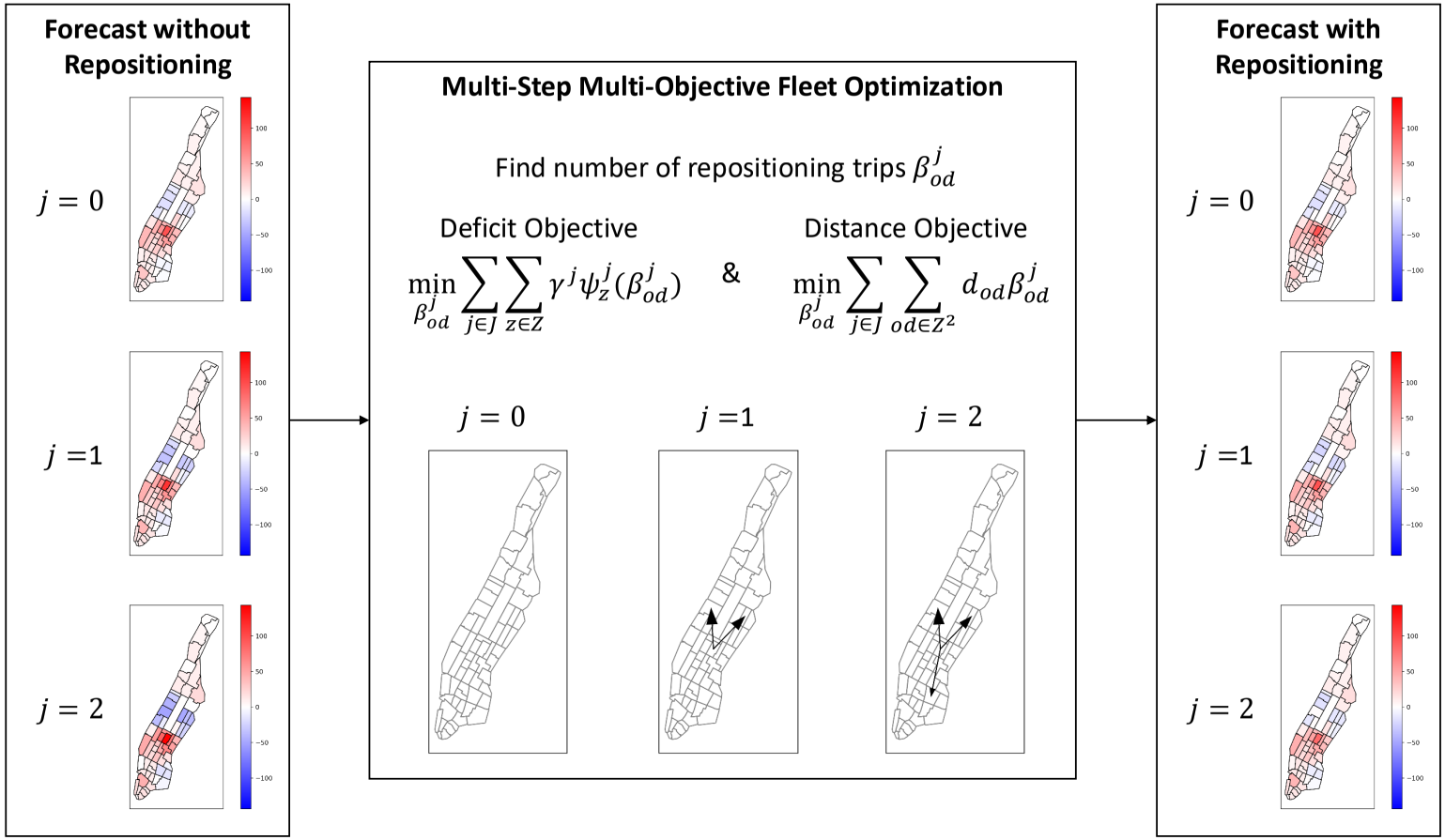


Figure 3.8: On the multi-step multi-objective repositioning problem. The illustrations show an arbitrary scenario and the colorbar represents the corrected imbalance per zone in this scenario.

There are several constraints that describe the dynamics of the multi-time-step system:

$$\text{s.t. } I_z^1 = \underbrace{|V_z^I| + |V_z^{P,1}| - \sum_{d \in Z} \Lambda_{zd}^1}_{\text{forecast without repositioning}} \underbrace{- \sum_{d \in Z} \beta_{zd}^1 + \sum_{o \in Z} \beta_{oz}^1 \delta_{oz}^{1,1}}_{\text{repositioning departures and arrivals}} \qquad \forall z \in Z \tag{3.39b}$$

$$I_z^j = \underbrace{Y_z^j}_{\text{stack}} + \underbrace{\sum_{o \in Z} \bar{\Lambda}_{oz}^j - \sum_{d \in Z} \Lambda_{zd}^j}_{\text{forecast}} \underbrace{- \sum_{d \in Z} \beta_{zd}^j + \sum_{o \in Z} \sum_{j'=1}^{j} \beta_{od}^{j'} \delta_{od}^{j',j}}_{\text{repositioning departures and arrivals}} \qquad \forall z \in Z \; \forall j \in J \setminus \{1\} \tag{3.39c}$$

$$\sum_{d \in Z} \beta_{zd}^j \leq Y_z^j \qquad \forall z \in Z \; \forall j \in J \tag{3.39d}$$

$$Y_z^1 = |V_z^I| \qquad \forall z \in Z \tag{3.39e}$$

$$Y_z^j = \max\left(I_z^{j-1}, 0\right) \qquad \forall z \in Z \; \forall j \in J \setminus \{1\} \tag{3.39f}$$

$$\bar{I}_z^j = \sum_{z' \in Z} K_{z,z'} \cdot I_{z'}^j \qquad \forall z \in Z \; \forall j \in J \tag{3.39g}$$

$$\psi_z^j = \max\left(-\bar{I}_z^j, 0\right) \qquad \forall z \in Z \; \forall j \in J \tag{3.39h}$$

$$\beta_{od}^j \in \mathbb{N}_0^+ \qquad \forall od \in Z^2 \; \forall j \in J \tag{3.39i}$$

Equations (3.39b) and (3.39c) describe the computation of the zone-imbalance considering possible vehicle repositioning decisions as described with equation (3.37). Equation (3.39d) constrains the number of vehicles repositioned out of a zone to be smaller than the vehicle stack variable Y_z^j at the beginning of the forecast step j. As described by equation (3.39f), the initial vehicle stack is determined by the number of idle vehicles at the time the repositioning algorithm is called. For all following forecast steps j, the stack is derived from the imbalance of the previous step $j-1$. If the imbalance I_z^{j-1} is positive, it can be expected that vehicles are idle at the beginning of step j. If on the other hand, a vehicle shortage is expected and the imbalance value is negative, the stack $Y_z^j = 0$. These two conditions can be summarized in the condition $Y_z^j = max\left(I_z^j, 0\right)$ in equation (3.39f). The reachability correlations between zones are applied in equation (3.39g). As explained earlier, equation (3.39h) defines the computation of vehicle deficits. Finally, equation (3.39i) is the integrality condition ensuring that an integer number of vehicles is repositioned.

Regardless of the approach to this multi-step multi-objective problem, the process after solving the respective problems is the same: The values for β_{od}^1 of the optimal solution determine the number of vehicles that should be repositioned at time t_i when the algorithm is called. The remaining part of the solution to this model-predictive control approach will

not be realized as the repositioning algorithm will be called again at time $t_i + \delta T^M$, which is the start of the forecast step $j = 2$, and recompute the optimal solution within the new time horizon. Finally, the zone-to-zone trips have to be transformed into explicit changes in vehicle plans. For this purpose, repositioning trips are assigned in random order, i.e., for each od-pair β^1_{od} idle vehicles in zone o are randomly chosen and receive a plan stop $X = (x_d, t^s = 0, t^l = t_i + t_{od}, t^e = t_i, R^+ = \emptyset, R^- = \emptyset, lock = True)$ at the centroid of zone d. The $lock$ flag signalizes that the stop is locked and that vehicle v cannot remove this stop from its plan or add any other stop before it. This is important as the short-term decisions do not consider the forecasts and thereby do not recognize the trip purpose, which is to decrease the vehicle deficit in another area in the near future.

The multi-objective multi-step problem (3.39) has several challenges:

1. even though the two objectives are already monetarized, it is not clear whether a simple sum approach is a good way to tackle the multi-objective problem
2. there are two $\max(.)$ expressions making the problem non-linear
3. it is a mixed integer problem

The multi-objective aspect is treated later, where different formulations are introduced. The second and third problems could be tackled by building a Markov Decision Process model (with multiple steps) and apply some dynamic programming methods to solve it; probably some approximations are necessary for larger scale problems. This thesis shows another approach, in which a modification of the problem formulation allows to solve all forecast steps at once in a single mixed integer linear program.

For this purpose, the $\max(.)$ terms need to be linearized. For the opportunity costs ψ, a relaxation of the $\max(.)$ constraint can be applied. Instead of equation (3.39h), two linear constraints can be introduced:

$$\psi^j_z \geq -\bar{I}^j_z \qquad \forall z \in Z \; \forall j \in J \qquad (3.40\text{a})$$

$$\psi^j_z \geq 0 \qquad \forall z \in Z \; \forall j \in J \qquad (3.40\text{b})$$

Equation (3.40b) ensures that there are no negative deficits. Equation (3.40a) states that the deficit is at least as large as the negative imbalance value. An optimization problem with this constraint allows deficits to become larger than the negative imbalance. However, as one objective is to minimize this deficit and the second objective has no benefits of larger ψ^j_z, the optimal solution will always satisfy equation (3.39h).

The same relaxation trick is not directly applicable to equation (3.39f), because there is no objective term trying to minimize the vehicle stacks Y^j_z in the current formulation. To overcome this problem, the set of repositioning actions is limited to ensure vehicle conservation with respect to zones in each forecast step. The only term breaking this vehicle conservation originates from $\delta^{j',j}_{od}$ as vehicles will not be counted to any zone if $j' \neq j$. Therefore, a zone-based vehicle conservation exists if repositioning trips ($\beta^j_{od} > 0$) are only allowed when they arrive in the respective destination zone in the same forecast step j. Assuming the existence of this zone-based vehicle conservation, $\sum_{z \in Z} Y^j_z - \psi^j_z$ is constant, i.e., independent of repositioning actions β^j_{od}, and an objective $\min \sum_{j \in J} \sum_{z \in Z} Y^j_z$ can be added without changing the general problem:

1. as long as vehicles are moved between zones with vehicle surplus, $\sum_{z \in Z} Y_z^j$ is constant
2. assume a vehicle is moved from a zone, where it would contribute to the stack, to a zone with deficit. Then the stack-objective is better and additionally, the respective deficit ψ_z^j is reduced, which reduces the deficit objective by the same amount

Hence, a minimization of the sum of deficits and the sum of stacks are aligned objectives and a replacement of objective is possible:

$$\min \sum_{j \in J} \sum_{z \in Z} \bar{f} \cdot \gamma^j \cdot \psi_z^j \rightarrow \min \sum_{j \in J} \sum_{z \in Z} \bar{f} \cdot \gamma^j \cdot \frac{1}{2} \left(\psi_z^j + Y_z^j \right) \tag{3.41}$$

With the minimization of Y_z^j as part of the objective, the linearization of $\max(.)$ term is possible again and equation (3.39f) can be replaced by

$$Y_z^j \geq I_z^{j-1} \qquad \forall z \in Z \; \forall j \in J \setminus \{1\} \tag{3.42a}$$

$$Y_z^j \geq 0 \qquad \forall z \in Z \; \forall j \in J \setminus \{1\} \tag{3.42b}$$

The limitation of repositioning trips to those ending in the same forecast time horizon, has further advantages. The first one is a reduction in the solution space. The optimization problem (3.39) contains a total of $|J| \cdot |Z|^2$ integer decision variables, which obviously is reduced by the limitation. Additionally, the operator can decide for vehicle v to make a very long repositioning trip at the first forecast step without the limitation. Due to the $lock$ status, this vehicle will not be available for any other assignments throughout its trip to the destination zone. Splitting the long trip into multiple short trips per forecast step has a similar effect on the imbalance and trip distance in the long run. However, the vehicle will become available on its trajectory for possible user assignments. Moreover, the decision to send the vehicle to the mid-term destination can be reassessed at each forecast step. If it is still meaningful, the original vehicle or another closeby vehicle will drive further towards the mid-term destination; if however the realized system state is other than expected, the operator can easily adapt its repositioning plan.

Therefore, all β_{od}^j with $\delta_{od}^{jj} \neq 1$ are set to 0 by adding the cuts

$$\beta_{od}^j - \delta_{od}^{jj} \leq 0 \quad \forall od \in Z^2 \; \forall j \in J \tag{3.43}$$

Alternatively, the unnecessary variables can be removed from the problem altogether by defining the sets of possible OD-pairs Z_{OD}^2, as well as the sets of possible destinations Z_z^d and origins Z_z^o for each zone z and replace Z and Z^2 in all sums and for-all statements in the optimization problem (3.39).

Even with the reduction in variables, the problem is still a complex and computationally expensive mixed integer linear program. A relaxation into a linear problem with subsequent quantization by rounding (down) of the $\boldsymbol{\beta}$ variables will be used as a computationally very efficient approximation of the optimal solution. The actions are determined from rounding down the part of the solution to the relaxed problem representing the repositioning actions in the first time horizon:

$$\beta_{od}^1 = \left\lfloor \bar{\beta}_{od}^1 \right\rfloor \tag{3.44}$$

In summary, the original problem is reformulated as

$$\min \sum_{j\in J}\sum_{z\in Z}\frac{1}{2}\cdot\bar{f}\cdot\gamma^j\cdot\psi_z^j \;\;,\;\; \min \sum_{j\in J}\sum_{z\in Z}\frac{1}{2}\cdot\bar{f}\cdot\gamma^j\cdot Y_z^j \;\;,\;\; \min \sum_{j\in J}\sum_{od\in Z_{OD}^2} c^D\cdot d_{od}\cdot\beta_{od}^j \quad (3.45a)$$

subject to:

$$I_z^1 = |V_z^I| + |V_z^{P,1}| - \sum_{d\in Z}\Lambda_{zd}^1 - \sum_{d\in Z_z^d}\bar{\beta}_{zd}^1 + \sum_{o\in Z_z^o}\bar{\beta}_{oz}^1 \qquad \forall z\in Z \qquad (3.45b)$$

$$I_z^j = Y_z^j + \sum_{o\in Z}\bar{\Lambda}_{oz}^j - \sum_{d\in Z}\Lambda_{zd}^j - \sum_{d\in Z_z^d}\bar{\beta}_{zd}^j + \sum_{o\in Z_z^o}\bar{\beta}_{oz}^j \qquad \forall z\in Z\ \forall j\in J\setminus\{1\} \qquad (3.45c)$$

$$\sum_{d\in Z}\bar{\beta}_{zd}^j \le Y_z^j \qquad \forall z\in Z\ \forall j\in J \qquad (3.45d)$$

$$Y_z^1 = |V_z^I| \qquad \forall z\in Z \qquad (3.45e)$$

$$Y_z^j \ge I_z^{j-1} \qquad \forall z\in Z\ \forall j\in J\setminus\{1\} \qquad (3.45f)$$

$$Y_z^j \ge 0 \qquad \forall z\in Z\ \forall j\in J\setminus\{1\} \qquad (3.45g)$$

$$\bar{I}_z^j = \sum_{z'\in Z} K_{z,z'}\cdot I_{z'}^j \qquad \forall z\in Z\ \forall j\in J \qquad (3.45h)$$

$$\psi_z^j \ge -\bar{I}_z^j \qquad \forall z\in Z\ \forall j\in J \qquad (3.45i)$$

$$\psi_z^j \ge 0 \qquad \forall z\in Z\ \forall j\in J \qquad (3.45j)$$

$$\bar{\beta}_{od}^j \ge 0 \qquad \forall od\in Z_{OD}^2\ \forall j\in J \qquad (3.45k)$$

The following paragraphs describe several approaches to address the multi-objective aspect of the problem.

Linear Weighted Sum (LWS) Strategy: Since the objectives are defined in monetized fashion, a simple sum approach is an obvious choice:

$$\min_{\boldsymbol{\psi},\boldsymbol{Y},\boldsymbol{\beta}} \sum_{j\in J}\left(\sum_{z\in Z}\frac{1}{2}\cdot\bar{f}\cdot\gamma^j\cdot\left(\psi_z^j + Y_z^j\right) + \sum_{od\in Z_{OD}^2} c^D\cdot d_{od}\cdot\beta_{od}^j\right) \qquad (3.46)$$

Varying the parameter $\gamma\in]0,1]$ shifts the relative importance of the opportunity cost objective to the distance costs related to the repositioning trips. This approach is called *Linear Weighted-Sum-[γ] Strategy*, where γ is replaced with the respective value used in the case study.

Linear Weighted Sum with Reduced Availability (LRA) Strategy: In the LWS problem formulation, repositioning vehicles are considered available at the destination zone as if they were already in stack at the beginning of the forecast step. The repositioning problem (3.45) can be extended with a minor adaption to implement and test the consideration of the amount of time a repositioning vehicle will be available in the new zone. This can be achieved by adjusting the computation of the imbalance $\bar{I}$ in equation $(3.45\mathrm{h})$ to

$$\bar{I}_z^j = \sum_{z'\in Z} K_{z,z'}\cdot\left(I_{z'}^j - \sum_{o\in Z_z^o}\beta_{oz}^z\cdot\left(1-\left(\Delta T^M - t_{oz}\right)\right)\right) \quad \forall z\in Z\ \forall j\in J \qquad (3.47)$$

This is denoted by *Linear Weighted Sum with Reduced Availability (LRA)* strategy. Note that the change is not made in the computation of the independent zone imbalances I (equations (3.45b) and (3.45b)) as vehicles should be fully counted at the end of the forecast step. If the change would be made on the imbalance level, the stack equation (3.45f) would require a modification. The idea of this reduced availability of repositioning vehicles is also used by Wallar, van der Zee, et al. [2018] but in a very different methodology. From a multi-objective point of view, this approach can be interpreted as reduced weights on the opportunity costs.

Linear Two-Step (LTS) Strategy: Another approach for the multi-objective problem is to focus as much as possible on improving service and reducing opportunity costs. This can be done with the lexicographic approach denoted by *Linear Two-Step (LTS)* strategy: the sums of deficits and stacks are minimized first and the solution creates constraints for the subsequent optimization of the distance objective.

This can be achieved by defining new variables $\boldsymbol{\Delta I}$

$$\Delta I_z^{j+} = \sum_{o \in Z_z^o} \beta_{oz}^j \qquad \forall z \in Z \; \forall j \in J \tag{3.48a}$$

$$\Delta I_z^{j-} = \sum_{d \in Z_z^d} \beta_{zd}^j \qquad \forall z \in Z \; \forall j \in J \tag{3.48b}$$

In Syed, Dandl, Kaltenhäuser, et al. [2021], the two step approach allowed a separation of the decision variables as it is possible to write constraints for ΔI_z^{j+} to guarantee a feasible solution for equation (3.48) in the second step. Unfortunately, this is not possible in the multi-step approach whose od-relations satisfy equation (3.43). Therefore, different from Syed, Dandl, Kaltenhäuser, et al. [2021], the od-specific variables β_{od}^j cannot be removed from the first optimization problem of the two-step approach.

Hence, the two-step approach first solves the optimization problem with constraints from (3.45) and the objective

$$\min_{\boldsymbol{\psi}, \boldsymbol{Y}, \boldsymbol{\beta}} \sum_{j \in J} \left(\sum_{z \in Z} \frac{1}{2} \cdot \bar{f} \cdot \gamma^j \cdot \left(\psi_z^j + Y_z^j \right) + c_m \cdot \sum_{od \in Z_{OD}^2} \cdot \beta_{od}^j \right) \tag{3.49}$$

with a small number $c_m \ll \bar{f}$ ensuring the priority of the opportunity costs.

Only the number of vehicles to be repositioned in the first forecast step $j = 1$ are relevant for the applied actions. After the computation of ΔI_z^{j+} and ΔI_z^{j-} for $j = 1$, the secondary objective optimization problem can be posed:

$$\min_{\boldsymbol{\beta}} \sum_{od \in Z_{OD}^2} c^D \cdot d_{od} \cdot \beta_{od}^1 \tag{3.50a}$$

$$\text{s.t.} \sum_{o \in Z_z^o} \beta_{oz}^1 = \left\lfloor \Delta I_z^{1+} \right\rfloor \qquad \forall z \in Z \tag{3.50b}$$

$$\sum_{d \in Z_z^d} \beta_{zd}^1 = \left\lfloor \Delta I_z^{1-} \right\rfloor \qquad \forall z \in Z \tag{3.50c}$$

$$\beta_{od}^1 \in \mathbb{N}^+ \qquad \forall od \in Z_{OD}^2 \tag{3.50d}$$

where $\lfloor . \rfloor$ operator denotes the floor-operator rounding down the values to the next integer value. As the constraint matrix is totally unimodular, this rounding guarantees that the solution of the relaxed linear problem with $\beta_{od} \in \mathbb{R}^+$ is integral. Formally, the second problem reflects the optimal rebalancing problem by M. Pavone et al. [2012]. Nevertheless, the methods are quite different in the determination of ΔI_z^{1+} and ΔI_z^{1-}.

Moreover, it should be noted that $\sum_{od \in Z_{OD}^2} \cdot \beta_{od}^j$ has to be part of the first-level objective. From a supply-demand balance point of view, all solutions with $\bar{\beta}_{od}^j - \bar{\beta}_{do}^j = \beta_{od}^j + \beta_{do}^j$ result in the same imbalance values. Nevertheless, if the first optimization problem would select the solution to e.g. send vehicles between two zones back and forth, these vehicles would have to be considered in the secondary problem (3.50), which is obviously not good.

Quadratic Two-Step (QTS) Strategy: Another promising modification is the utilization of a quadratic imbalance objective [Syed, Dandl, Kaltenhäuser, et al., 2021]. A quadratic formulation puts more weight on zones with high vehicle deficit or surplus. Hence, idle vehicles from zones with large surplus are more likely to be repositioned than vehicles from zones with minor surplus and zones with high expected deficit are more likely to receive vehicles than zones with small deficits.

Since the quadratic objective cannot be monetized easily and the results in [Syed, Dandl, Kaltenhäuser, et al., 2021] showed good results for the lexicographic method, this quadratic objective is used in conjunction with the two-step approach with the secondary problem described by equations (3.50). The primary optimization has constraints from (3.45) and the objective

$$\min_{\boldsymbol{\psi}, \boldsymbol{Y}, \boldsymbol{\beta}} \sum_{j \in J} \left(\sum_{od \in Z_{OD}^2} \left(\psi_z^j\right)^2 + \left(Y_z^j\right)^2 + c_m \cdot \sum_{od \in Z_{OD}^2} \cdot \beta_{od}^j \right) \tag{3.51}$$

where $c_m \ll 1$ ensures the priority of the imbalance terms. This method is denoted by *Quadratic Two-Step (QTS)* strategy in this thesis.

Quadratic problems with linear constraints can become computational intractable. They are either NP-hard [Pardalos and Vavasis, 1991] or solvable in polynomial time [Kozlov et al., 1980] if and only if the matrix $\boldsymbol{Q}$ in the bilinear form $\boldsymbol{x}^T Q \boldsymbol{x}$ representing the quadratic variables is positive definite. Here, x represents the vector of all ψ_z^j and Y_z^j and Q is the unit matrix (of dimension $2 \cdot |Z| \cdot |J|$), which is positive definite. Therefore, the quadratic problem can be solved efficiently.

After ΔI_z^{1+} and ΔI_z^{1-} are determined from the solution, problem (3.50) is solved to find the assignments with minimum expected distance.

Quadratic Deficit and Linear Stack Two-Step (QDLSTS) Strategy The quadratic formulation in equation (3.51) also generates repositioning flows between zones with vehicle surplus to balance the surplus. This does not reflect the original idea anymore, where repositioning should be used to counteract vehicle deficits. To avoid this effect but still prioritize zones with large deficits, one last problem formulation is created in this thesis, in which the deficits are quadratic variables, the vehicle stacks remain linear and the distance objective is once again treated in a two-step approach with the optimization problem (3.50). The primary

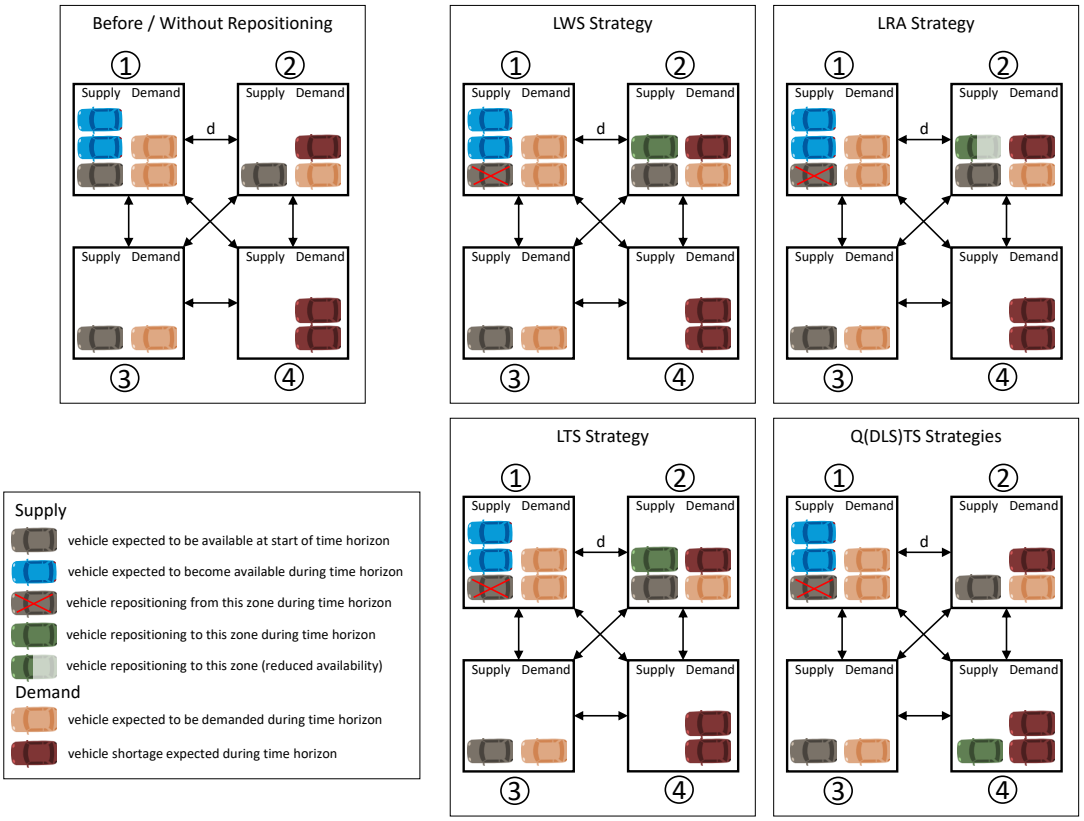


Figure 3.9: Comparison of one forecast step for the described repositioning strategies for an example with 4 independent zones ($\bar{\boldsymbol{I}} = \boldsymbol{I}$) in case the expected profit from serving an additional user outweighs the costs of repositioning between zones.

problem contains the constraints from the original problem (3.45) and the objective

$$\min_{\boldsymbol{\psi},\boldsymbol{Y},\boldsymbol{\beta}} \sum_{j\in J} \left(\sum_{od\in Z^2_{OD}} \frac{1}{N^{esq}} \left(\psi^j_z\right)^2 + Y^j_z + c_m \cdot \sum_{od\in Z^2_{OD}} \cdot \beta^j_{od} \right) \tag{3.52}$$

where $c_m \ll 1$ ensures the priority of the imbalance terms and N^{esq} determines the size of the deficit, from which deficit value ψ^j_z on the quadratic objective should be valued higher than it was the case in the linear problem formulation, i.e., $\psi^j_z / N^{esq} > 1$.

As before, the matrix Q of the bilinear form is a unit matrix (this time of dimension $|Z|\cdot|J|$), thereby ensuring computational efficiency.

Again, after ΔI^{1+}_z and ΔI^{1-}_z are determined, problem (3.50) is solved to find the assignments with minimum expected distance.

Illustrative Example for Application of Different Repositioning Strategies Figure 3.9 illustrates different actions based on the repositioning strategies for a simple example with one forecast step. In four independent zones ($\bar{\boldsymbol{I}} = \boldsymbol{I}$), fares and distance costs are chosen such that serving an additional requests creates more revenue than costs. In this example, the

LWS, LRA and LTS strategies determine to send one vehicle from zone 1 to the nearby zone 2 (distance d). For the RA strategy, the vehicle is not counted as fully available in zone 2. In contrast, the QTS and QDLSTS strategies favor to drive a little more distance ($\sqrt{2} \cdot d$) to reduce the larger imbalance. If the ratio of expected fare over costs of driving in the example would change in the favor of costs, the LS and RA strategy would not reposition anymore, while the two-step strategies are not sensitive to this ratio thereby making the same decisions as illustrated.

This illustrative example does not make any statements about the final fleet performance, which has to be evaluated by numerical simulations. In the operational case study (section 3.6), the newly developed density-based repositioning strategies are compared with each other and an existing strategy from literature. Therefore, the real-time rebalancing strategy developed by M. Pavone et al. [2012], which is widely used in transportation research due to its simplicity and effectiveness, serves as a baseline. The algorithm is briefly described in Appendix B.3.

3.4.3 Dynamic Pricing

All previously discussed repositioning strategies aim to bring vehicles where they are needed, and best ahead of time. In other words, repositioning tries to match supply to demand. However, in times of very high demand, it might not be possible to equalize all vehicle shortages. In the illustrated example (Figure 3.9), such situation occurs and not all shortages can be removed. In these situations, dynamic pricing can be utilized to reduce demand in order to match demand to supply.

The fare system in equation (3.1) considers both a minimum base fare f^B and a distance-dependent fare f^D. In a transportation system, the relation of both components can influence whether the service serves shorter or longer trips [Wilkes et al., 2021]. In theory, dynamic pricing can scale both differently. However, it is not instinctively clear that scaling one over the other brings benefits to the system. The modeling of price sensitivity to both components requires complicated demand models and data that is not available to the author. Hence, a simpler approach, in which the total fare is scaled and users are sensitive to the total price, is chosen. In the simulation model, the operators can utilize time- and location-dependent dynamic pricing. In general, the modified fare for a request r reads

$$f_r = \beta^p_{od,t} \cdot \min\left(f^B, f^D \cdot d_r\right) \tag{3.53}$$

This thesis makes the simplifying assumption that the operator only applies dynamic pricing to match demand to supply when necessary, i.e., the operator sets fare scale factors $\beta^p_{od,t} \geq 1$ to decrease demand. In theory, it could also be possible that an operator reduces fares $\beta^p_{od,t} < 1$ in certain zones to bring surplus vehicles out of these zones. From an operator perspective, this can be advantageous as vehicles might end up where it would reposition it anyway but there is no guarantee. The future vehicle stacks in the repositioning strategy would have to be adapted to the expected changes, which would require a more complex interaction of repositioning and dynamic pricing strategies. Moreover, from a transportation system point of view, one should also consider from which modes this induced demand comes from, which is not in the scope of this thesis.

It is assumed that the fares should not change too frequently, which might be confusing and unwanted by AMoD users. Therefore, the time-dependency is determined by the mid-term

decision time step Δt^M. Every Δt^M, the operator determines a set of dynamic pricing factors $\beta^p_{od,t}$, which are valid for Δt^M when the next iteration of the pricing algorithm determines the pricing factor for the next time interval.

In the following paragraphs two dynamic pricing models are introduced. The first strategy aggregates all spatial information by utilizing the total fleet utilization as control variable, the second strategy uses the density-based forecast model.

Utilization-Based Dynamic Pricing This strategy assumes a relation between fleet utilization η, which is defined as the share of driving vehicles (regardless whether it has customers on board or not) and the probability that a new user request can be matched successfully: vehicles that are already in use have a smaller chance of accommodating the new user. Therefore, the dynamic pricing factor is modeled as a function of share η_t of utilized vehicles at time t. For simplicity, a piece-wise linear function is applied

$$\beta^p_t = \begin{cases} 1 & \eta_t < 0.75 \\ 1 + \beta^{PU} \cdot (\eta_t - 0.75) & \eta_t \geq 0.75 \end{cases} \tag{3.54}$$

where β^{PU} determines the maximum pricing factor in case of full fleet utilization.

One-Step Forecast Based Dynamic Pricing A more complex model utilizes information about the price sensitivity of users, the expected demand during the forecast horizons and the spatio-temporal information, in which zones $z \in Z$ vehicle deficits are expected.

The framework assumes that vehicle shortages in the first forecast step lead to unserved requests leaving the system. With the premise of this dynamic pricing approach to setting the fare at a level that demand matches the available supply, the number of fulfilled trips would be similar. Therefore, the forecasts of future steps in the multi-step approach should remain approximately the same, even with dynamic pricing ($\beta^p \geq 1$) in the current step. Therefore, the repositioning strategy can determine its actions first and this dynamic pricing strategy is called after the repositioning strategy is performed. It uses the same forecast horizon step ΔT^M, but only one step is required. Moreover, all repositioning actions can already be considered in the computation of the respective deficits $\bar{I}^1_z$.

It is assumed that the price sensitivity of demand can be modeled by a time- and space-invariant function D such that the price-sensitive demand $\Lambda_z(\beta_z)$ during the next forecast step ($j = 1$) is given by

$$\Lambda_z(\beta^p_z) = D(\beta^p_z) \cdot \Lambda^1_z \tag{3.55}$$

for all zones $z \in Z$, where Λ^1_z denotes the expected demand if the fare would not be adapted, i.e., for $\beta^p_z = 1$.

Let $\bar{I}^{1'}_z$ be the expected imbalance in zone z (considering reachability) for the upcoming forecast horizon after the repositioning actions are considered. Moreover, let $\bar{I}^{1-}_z = \min(\bar{I}^{1'}_z, 0)$ be the negative value of the expected remaining vehicle deficiency. Then the *One-Step Forecast Based Dynamic Pricing* strategy chooses the dynamic pricing factors such that

$$\Lambda_z(\beta^p_z) = \Lambda^1_z + \bar{I}^{1-}_z \tag{3.56}$$

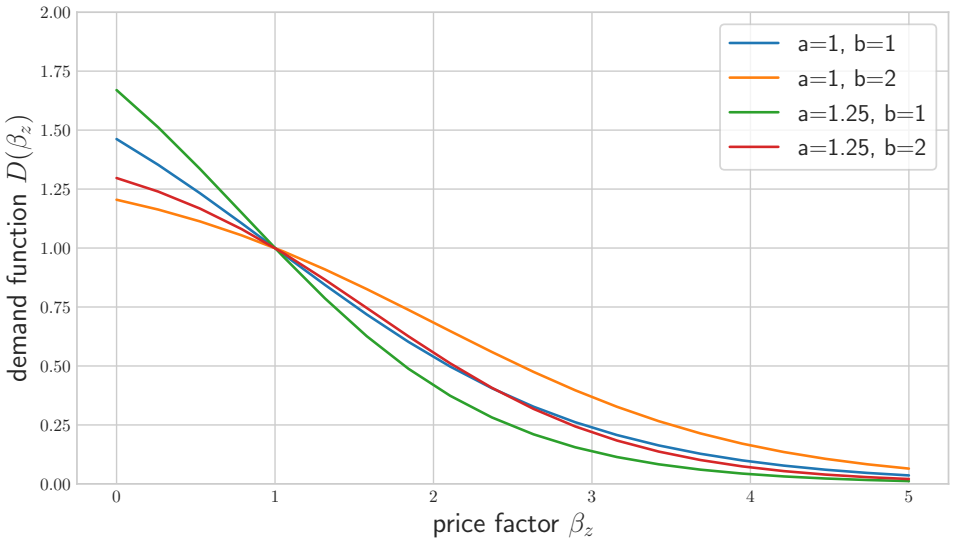


Figure 3.10: Examples for demand functions of form (3.59) representing the price sensitivity of AMoD users. The functions are normalized to return the expected demand (i.e., $D = 1$) for price factor 1.

With equation (3.55), it can be derived that

$$D(\beta_z^p) = 1 + \frac{\bar{I}_z^{1-}}{\Lambda_z^1} \tag{3.57}$$

It has to be noted that equation (3.56) can result in a negative value for $\Lambda_z(\beta_z^p)$ because of the impact of demand in neighboring zones on the reachability-corrected imbalance. Since a non-positive demand does not make any sense, $\hat{I}_z^{1-} = \max(\bar{I}_z^{1-}, \epsilon - \Lambda_z^1)$ with a small number $\epsilon > 0$ is defined.

The higher the price factor, the less people want to use the service. Hence, any normal demand function $D : \mathbb{R}^+ \to \mathbb{R}^+$ is strictly monotonous decreasing. Therefore, an inverse function D^{-1} exists:

$$\beta_z^p = D^{-1}\left(1 + \frac{\hat{I}_z^{1-}}{\Lambda_z^1}\right) \tag{3.58}$$

Figure 3.10 shows examples for demand functions of the form

$$D(\beta^p) = \frac{1 + \exp(a - b)}{1 + \exp(a \cdot \beta^p - b)} \tag{3.59}$$

where the enumerator is chosen to normalize $D(1) = 1$ (consistency of expected demand for normal fares). This functional form can be used to describe price sensitivities according to Talluri and van Ryzin [2004]. As an application in the transport sector, Hardt and Bogenberger [2021] describe the price-sensitivity of carsharing. Since uncertainties in this function are not in the scope of this thesis, it is assumed that equation (3.59) describes both the expected as well as the realized price sensitivity in the case study. With this definition, β_z^p

can be computed analytically

$$\beta_z^p = \frac{b}{a} + \frac{1}{a} \ln \left(\frac{1 + \exp(a - b)}{1 + (\hat{I}_z^{1-} / \Lambda_z^1)} - 1 \right) \tag{3.60}$$

A solution $\beta_z^p > 1$ exists for equation (3.60) for all zones $z \in Z$ with $\bar{I}_z^{1-} < 0$. For zones with $\bar{I}_z^{1-} = 0$, i.e., vehicle surplus or balance, $\beta_z^p = 1$. The solution does not have an upper bound (for arbitrarily small values of ϵ) and very large factors $\beta_z^p \gg 1$ would be the result of severe vehicle shortages. In order to keep fares within a certain range, it is assumed that the operator sets a threshold $\tilde{\beta}^p$, i.e.

$$\beta_z^p = \min \left[\tilde{\beta}^p, \frac{b}{a} + \frac{1}{a} \ln \left(\frac{1 + \exp(a - b)}{1 + (\bar{I}_z^{1-} / \Lambda_z^1)} - 1 \right) \right] \tag{3.61}$$

The pricing factors $\boldsymbol{\beta^p}$ are valid until the next mid-term step and are applied on the fares f_r of all requests in this time interval:

$$f_r = \beta_{z_r}^p \cdot \min \left(f^B, f^D \cdot d_r \right) \tag{3.62}$$

where z_r is the zone of the start location of request r.

3.5 Planning Long-Term Decisions

While short-term and mid-term decisions are repeatedly made within the evaluation period T based on the respective strategies, long-term decisions are made at the beginning and remain constant throughout the evaluation period T. Equation $(3.6\mathrm{c})$ essentially states that these parameters should be chosen such that the expected profit over the complete evaluation period is maximal. This expectation value can be determined by simulations.

The operator has to decide on an operating area Ω, a fleet size $|V|$, and values for the base and distance fares f^B and f^D. The profit function P behaves nicely with respect to fleet size and fares. For meaningful fare and cost structures, the profit function $P(|V|)$ is a convex function of fleet size $|V|$: starting from low vehicle numbers, the revenue and profit increase with increasing fleet size. At some optimal point, the AMoD demand starts to saturate and the marginal increase in revenue becomes smaller than the additional fix cost for increasing the fleet size. With price-sensitive demand and a fixed fleet size, the profit function $P(f^B, f^D)$ increases with the fare (there is zero revenue for $f^B = f^D = 0$). On the other end, for very large values of (f^B, f^D), almost nobody will use the AMoD service, which results in near-zero revenue as well. Since the transition can be expected to be rather smooth, $P(f^B, f^D)$ is convex and has a maximum.

For the combined multi-dimensional optimization, a scenario-based grid search or more advanced methods like gradient descent can be used. However, the evaluation of a data point in the $((|V|, f^B, f^D)$ space requires simulations of the AMoD system for the evaluation period T, which is computationally expensive. Hence, heuristic macroscopic methods to limit the search space for simulations can be useful.

If a demand estimation and a distribution of trip durations exists, a very simple method to estimate a lower bound for the required hailing fleet size is to compute the expected number of customers that are in the system at the same time:

$$|V|_{<} = \max_{t \in 0 \dots N^T} \left[\sum_{r \in R} H\left(t - \tau_r\right) \cdot H\left(\tau_r + t[x_r^o \rightarrow x_r^d] - t\right) \right] \tag{3.63}$$

where $H(.)$ is the heaviside step function. The first heaviside function is non-zero after the request time and the second one is non-zero before the expected arrival time if the requests started at the request time. Hence, the combination of both approximates the time a request is in the system. The lower bound derived from equation (3.63) does not yet consider that additional vehicles will be needed for the empty pick-up trips. Based on literature (e.g., [M. Hyland and Mahmassani, 2018a]), another $10 - 30$ % of vehicles should be added. For pooling, this thesis uses a simple approach and approximates $|V|_{<}$ for pooling by dividing the respective number of vehicles for hailing by estimated sharing factors.

An estimation of demand as function of the base fare and distance fare is quite difficult and depends on a lot of factors such as competing AMoD services, their behavior, private vehicle ownership, availability and pricing of line-based PT services, and the attitudes of the population to name a few of them. Therefore, the scenario-based evaluation will be necessary. Service design considerations like app-design, interaction of users and operator and level-of-service targets for average waiting and detour time also affect the demand.

Even though the operating area could be treated as a continuous variable, AMoD operators are likely to use existing spatial boundaries to make the operating area comprehensible for users. Hence, the optimization will be the selection of the best scenario from a discrete set of possible operating areas.

3.6 Case Study: Setup

In this section, the operational strategies are evaluated in a case study for the city of Munich, Germany. In order to address the stochastic dynamic problem, the evaluation will be based on simulations. Hence, the agent-based simulation model FleetPy[11] is introduced next. Subsequently, the case study setting, i.e., the data input of this simulation model, is described for the city of Munich. As this thesis introduces several operational strategies, the evaluation is split into different studies analyzing certain aspects, which will be described in section 3.6.3 about the scenario setup. After presenting and discussing the results of this studies, some findings that will be the basis for chapter 4 about regulating AMoD systems will be summarized in the conclusion.

3.6.1 Agent-Based Simulation Model

A simulation model is designed to capture the dynamics of the AMoD system, i.e., states S_i, actions A_i, exogenous variables s_i, state transitions Θ and rewards Ψ in the original problem

[11] See Appendix B.5 for some more general information about FleetPy.

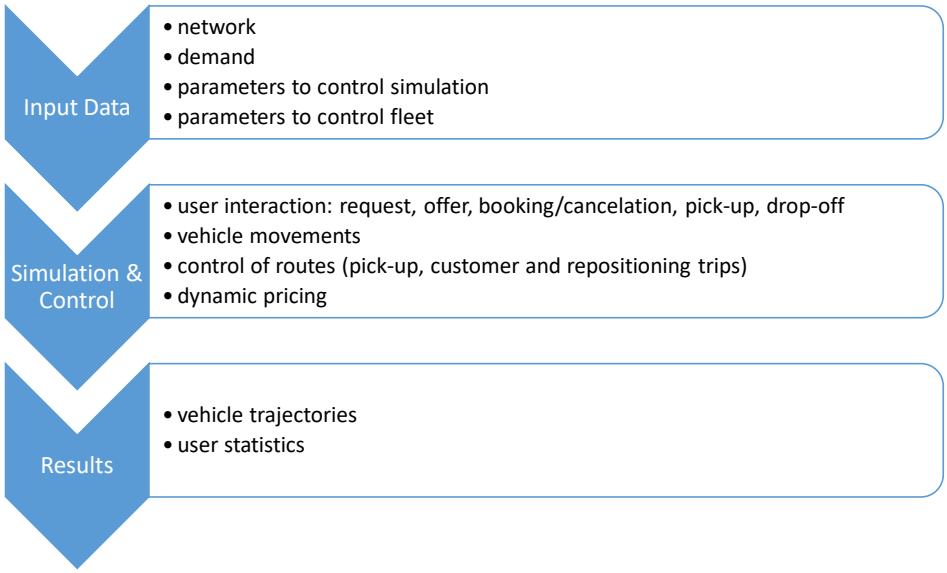


Figure 3.11: Phases of a single agent-based AMoD simulation.

formulation (3.4). It contains several agent types. Each user request r is represented by an agent, which contains the information $r = (x_r^p, x_r^d, \tau_r, \tau_r^e, \tau_r^l, \tau_r^d, f_r^{max})$. Each AMoD vehicle v is explicitly modeled by $v = (x_v, R_v^o, X_v^0)$, i.e., its current position, the current on-board requests, and the route to the next plan stop and its respective task describing, for instance, which customers should board/disembark the vehicle there. X_v can also be empty if a vehicle is idle. Finally, the operator is modeled by an agent, which centrally controls all AMoD vehicles.

The simulation can be divided into several phases. As displayed in Figure 3.11, the *Initialization Phase* processes the input data that define a simulation scenario, the *Simulation Step Phase* performs the time-controlled simulation steps which applies the fleet control strategies to the original problem given by equation (3.4), and an *Evaluation Phase*, where the recorded vehicle trajectory and user statistic results are evaluated.

Initialization Phase: The agent-based simulation model requires network data, i.e., a network graph $G = (N, E)$ and the attributes distance d_e and travel time t_e for all edges $e \in E$. Even though the framework is capable of dynamic travel times $t_e(t)$, the operational case study assumes constant values, which allows the creation of a constant node-to-node travel time matrix. Travel costs — both distance and travel time — can be read directly from this matrix, and routes can be determined very efficiently with the algorithm described in Appendix B.2. Positions in the network are described by tuples (n_0, n_1, rel), where n_0 and n_1 define the edge a vehicle is on and rel determines the relative position on the edge. If a position is exactly on a node, only n_0 is set, n_1 is not defined and rel is 0.

The total demand, which represents the amount of users making requests to the AMoD system, is exogenous and an upper bound for the number of served customers. In contrast, the actually served demand for the AMoD system is endogenous as users requesting service can leave the system without traveling with AVs of the AMoD operator.

The user model is determined by simulation parameters. It is assumed that level-of-service

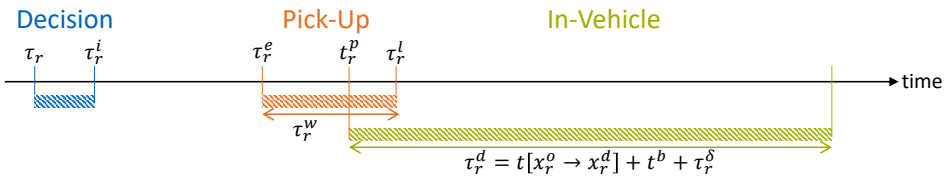


Figure 3.12: On the time constraint definition in this thesis. The pick-up has to take place within the interval $[\tau_r^e, \tau_r^l]$ and the detour parameter determines the latest time of arrival at the drop-off location x_r^d at the time of the pick-up t_r^p.

limits, namely the duration of the maximum waiting time window τ_r^w and the relative detour $\tilde{\tau}_r^\delta$, are set and accepted by both operator and users. Moreover, this case study assumes that both parties make instantaneous decisions such that the latest decision time τ_r^i of requests is equal to their request time ($\tau_r^i = \tau_r$). While the simulation framework is capable of non-instantaneous decisions, $\tau_r^i = \tau_r$ will be selected for the operational case study as it will become important for chapter 4. In most scenarios, the users want immediate service ($\tau_r^e = \tau_r$) and τ^w becomes a maximum waiting time. The relative detour parameter $\tilde{\tau}_r^\delta$ is defined by

$$\tilde{\tau}_r^\delta = \frac{\tau_r^d}{t[x_r^o \rightarrow x_r^d] + t^b} - 1 = \frac{\tau_r^\delta}{t[x_r^o \rightarrow x_r^d] + t^b} \tag{3.64}$$

where τ_r^δ is the absolute detour. The simulation framework records a user as in-vehicle from the beginning of its boarding process to the beginning of its disembarking process. Hence, the boarding time t^b, another simulation parameter, has to be added to the direct route time $t[x_r^o \rightarrow x_r^d]$ for the computation of the absolute and relative detour. Therefore, the latest arrival time is generated dynamically during the simulation at the time of the pick-up t_r^p (see Figure 3.12).

The simulation requires the evaluation period T as well as the simulation time step Δt as input. All requests $r \in R$ are set up in the initialization phase and put in the set of not revealed requests R^u that are unknown to the operator. A request r is defined by its attributes $r = (x_r^o, x_r^d, \rho_r, \tau_r, \tau_r^e, \tau^w, \tilde{\tau}_r^\delta)$.

For the scenarios with dynamic pricing, the sensitivity towards dynamic pricing is modeled. Therefore, the standard fare $f_r^{def} = \min\left(f^B, f^D \cdot d_r\right)$ is evaluated based on the distance of the fastest path. Next, the maximal acceptable fare is determined by drawing an acceptable dynamic pricing factor $\beta_r^{p,max} = f_r^{max}/f_r^{def}$ for each user such that the overall distribution reflects the demand curve $D(\beta^p)$ from equation (3.59). In this way, the operator does not know the exact price sensitivity of each user. The necessary distribution for $\beta_r^{p,max}$ is generated by the normalized function $-D'(\beta^p)$ representing the negative derivative of $D(\beta^p)$.[12]

Finally, all operator related parameters have to be initialized in order to define the fleet control. These parameters select the short-term and mid-term strategies and set the hyper-parameters of these strategies as well as the long-term variables for a single simulation. Furthermore, the initial vehicle distribution has to be determined for a simulation.

[12]Reasoning: given a certain pricing factor β_0^p, a variation of this price by $d\beta^p$ reduces the amount of demand by $\frac{dD(\beta^p)}{d\beta^p}\Big|_{\beta_0^p}$. Hence for a certain set of users, the number of users with acceptable pricing factor β_0^p is given by $-D'(\beta^p)$.

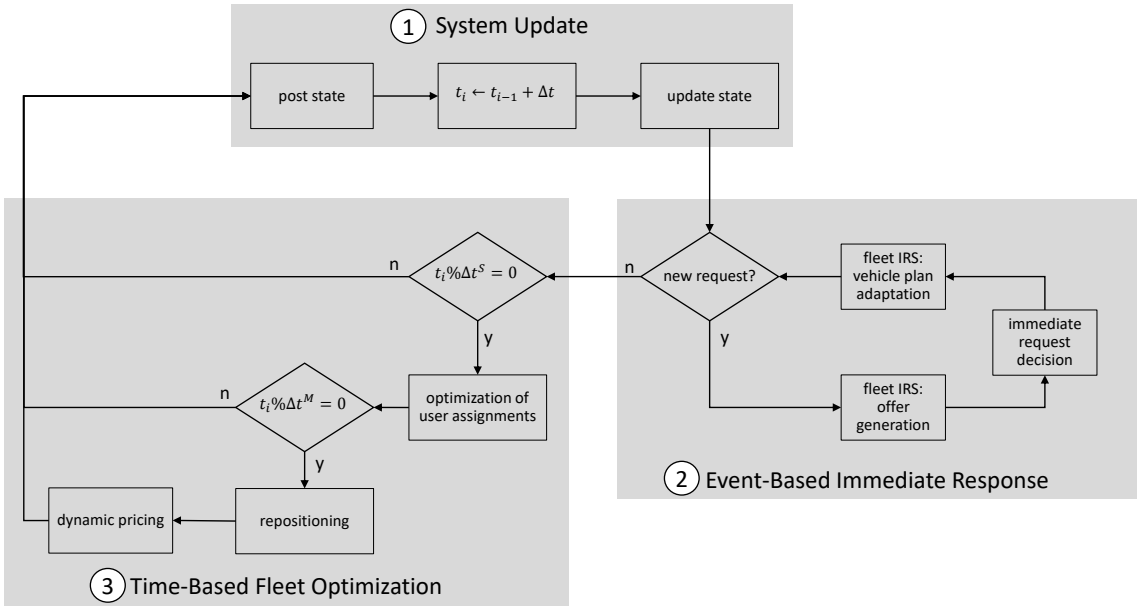


Figure 3.13: Flowchart of simulation step: (i) the system state is updated, (ii) an event-based immediate response model determines the booking behavior of requests, and (iii) the operator applies its fleet optimization strategies with certain periodicity.

Simulation Step Phase: Once the simulation is set up, a loop over time controls the simulation. In every time step i, the processes displayed in Figure 3.13 are performed.

(1): The *system update* denotes the state transition of the state

$$S_i = \left((x_v, R_v^o, \xi_v)_{v \in V}, R^a, \mathbf{\Lambda}, \bar{\mathbf{\Lambda}} \right) \tag{3.65}$$

from time t_i to time t_{i+1} as described by the state transition operator Θ: according to the current vehicle plans, each vehicle changes its position, picks up or drops off customers (thereby modifying their request status), and new requests become active. Moreover, stochastic information about expected future requests $\mathbf{\Lambda}$ and vehicle arrivals $\bar{\mathbf{\Lambda}}$ is updated. However, with the separation of scales, it is sufficient to update them every Δt^M since they are only used by the mid-term strategies.

(2): The *event-based immediate response* processes model the interaction of new user requests with the operator. Depending on the use case (hailing or pooling), the respective insertion heuristic creates an offer for every new request. These immediately accept or decline the offer depending on the traveler model and the offer. The requests that accept and decline the offer, are added to the request sets R^a and R^l, respectively.

(3): The operator can re-optimize the vehicle plans ξ_v. The processes in the *time-based fleet optimization* phase describe the strategies to determine actions based on user-assignment re-optimization, repositioning and dynamic pricing. These actions do not require any customer interaction and modify only the state variables ξ_v. The different strategies are not applied in every time step, but periodically with intervals in the short-term (Δt^S) or mid-term (Δt^M) time scales. The fleet control algorithms are implemented in a modular way in order to allow

different combinations of assignment, repositioning, and dynamic pricing strategies (described in section 3.3 and 3.4) to be applied in a single simulation.

To close the loop, the updated vehicle plans ξ_v and request sets R^a and R^l determine the post state S^+, according to which the system transitions to the next time step.

In order to have a tractable data log, not all endogenous changes of the system state are recorded. It is also not feasible to keep track of all actions, i.e., changes in vehicle plans. Instead of saving the coordinates (or network position) traces of thousands of vehicles driving hundreds of kilometers in a single day, only the stops and the sequence of nodes between these stops are recorded. A stop is also recorded whenever all tasks of a vehicle are assigned to other vehicles and the vehicle becomes idle. On the request side, the number of state changes remains feasible, therefore it is possible to record all of them. When a user accepts an operator offer, the fare is recorded for the respective user. In accordance with the definition of the time constraints, boarding and alighting are recorded when a user starts the respective boarding processes.

After the simulation time t_i reaches the end of the evaluation horizon T, no more requests enter the system. The remaining vehicle plans are completed in order such that users requesting a trip shortly before the end of the evaluation period are still recorded properly.

Evaluation Phase: The evaluation of the global operator objective

$$P = \sum_{r \in R^s} f_r - c^F \cdot |V| + \sum_{v \in V} c^D \cdot d_v \tag{3.66}$$

is straightforward from the recorded vehicle trajectories and user statistics. The fares can be read from the user statistics and the driven distance can be derived as all nodes that a vehicle passed on its trajectory are recorded. Additionally, key performance indicator (KPI)s from operator and user perspective can be defined.

From the operator perspective, the following KPIs, which can help to identify potential for improvement of the fleet, are automatically evaluated at the end of a simulation:

$$\text{mean fleet utilization} = \frac{1}{|V|} \sum_{v \in V} \frac{T - T_v^{idle}}{T} \tag{3.67a}$$

$$\text{peak fleet utilization} = \max_j \frac{1}{|V|} \sum_{v \in V} \frac{T^{eval} - T_{v,j}^{idle,eval}}{T^{eval}} \tag{3.67b}$$

$$\text{total VKT} = \sum_{v \in V} d_v \tag{3.67c}$$

$$\text{share of empty VKT} = \frac{\sum_{v \in V} d_v^{empty}}{\text{total VKT}} \tag{3.67d}$$

$$\text{share of repositioning VKT} = \frac{\sum_{v \in V} d_v^{repo}}{\text{total VKT}} \tag{3.67e}$$

$$\text{total PKT} = \sum_{v \in V} \sum_{\rho} \rho \cdot d_v^{\rho} \tag{3.67f}$$

$$\text{km-weighted average occupancy} = \frac{\text{total PKT}}{\text{total VKT}} \tag{3.67g}$$

$$\text{effective PKT} = \sum_{r \in R^s} d_r \tag{3.67h}$$

$$\text{relative saved distance} = \frac{\text{effective PKT} - \text{total VKT}}{\text{effective PKT}} \tag{3.67i}$$

The fleet utilization gives insight about the temporal usage of vehicles. In general, an operator wants to avoid long total idle times T_v^{idle} of vehicles as they only generate fixed costs but no revenues. On the other end of the spectrum, a fleet utilization near 100% can also be disadvantageous as there are no more reserves for additional requests. As demand is typically not constant over time, the peak utilization is a better indicator to check whether the fleet is at its limits. For the evaluation of the peak, the evaluation period T is divided into shorter evaluation intervals (with index j) of duration T^{eval} and the idle times are measured during the respective intervals. A large part of the total VKT is dictated by the number of requests and their desired trips, especially for hailing systems. The operator has full control over the empty pick-up and repositioning trips, which is why the share of empty and repositioning VKT are interesting quantities. For pooling systems, the relation between person kilometers traveled (PKT) and VKT is of interest. In this thesis, the pooling efficiency is measured by the km-weighted average occupancy and the relative saved distance. For the computation of the (km-weighted) average occupancy, each driven km is weighted by its respective occupancy ρ. This is an intuitive quantity but could in theory be inflated by taking long detours with multiple passengers on board. Such inflation is not possible for the relative saved distance indicator, which uses the effective PKT as basis for comparison. The effective PKT are defined as the sum of the trip distances assuming all served users would have taken the fastest path with a private vehicle.

Additionally, the following user perspective KPIs are evaluated:

$$\text{share of served requests} = \frac{|R^s|}{|R|} \tag{3.68a}$$

$$\text{mean waiting time} = \frac{1}{|R^s|} \sum_{r \in R^s} (t_r^p - \tau_r^e) \tag{3.68b}$$

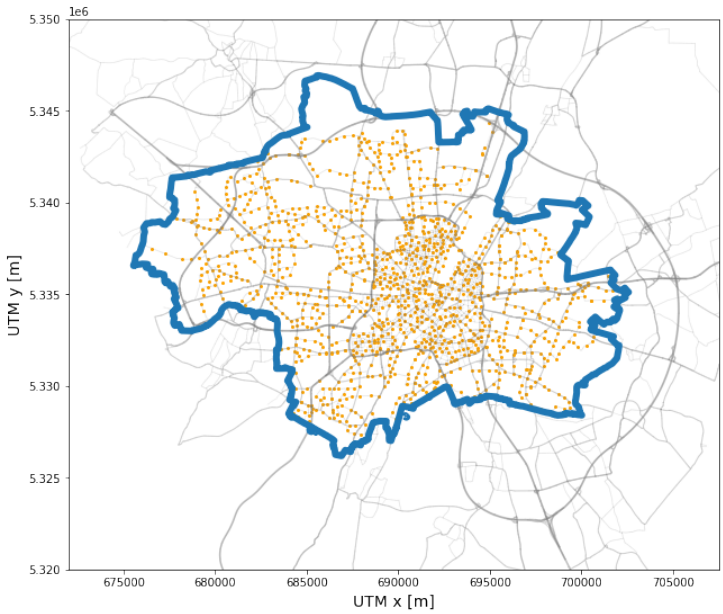
$$\text{mean travel duration} = \frac{1}{|R^s|} \sum_{r \in R^s} \left(t_r^d - t_r^p\right) \tag{3.68c}$$

$$\text{mean detour time} = \frac{1}{|R^s|} \sum_{r \in R^s} t_r^\delta = \frac{1}{|R^s|} \sum_{r \in R^s} \left(t_r^d - (t_r^p + t[x_r^o \to x_r^d] + t^b)\right) \tag{3.68d}$$

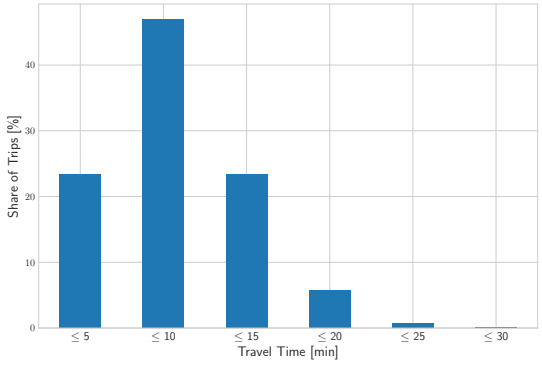
$$\text{mean relative detour time} = \frac{\text{mean detour time}}{\text{mean travel duration}} \tag{3.68e}$$

$$\text{mean fare per (direct) km} = \sum_{r \in R^s} \frac{f_r}{d_r} \tag{3.68f}$$

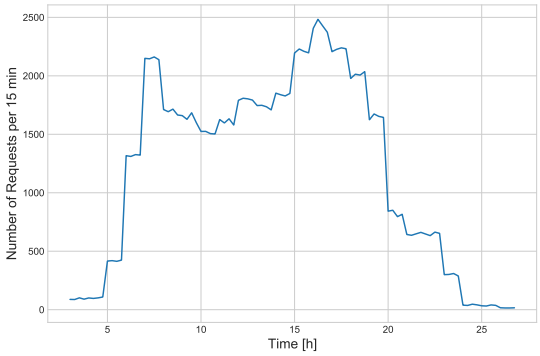
Since there are constraints for waiting and travel time and the maximum acceptable fare, the most significant quantity is the share of served requests, which shows for how many requests the operator is capable of satisfying these constraints.



(a) Municipality of Munich, its street network, and AMoD pick-up and drop-off points in UTM32 (EPSG:32632) coordinate system. Streets with more important functional road category are drawn with higher line width.



(b) Distribution of AMoD trip durations.



(c) AMoD trip rate over time in 5 % scenario.

Figure 3.14: Information on network and demand data of the case study.

3.6.2 Input Data and Data Pre-Processing

The case study is set up in Munich, Germany. The original operating area is determined by the official boundaries of the municipality. These are shown together with the street network extracted from a microscopic traffic model [Dandl, Bracher, et al., 2017] in Fig. 3.14a in the UTM32N (EPSG:32632) coordinate system, which will be used throughout this thesis. Moreover, this figure illustrates the street network access nodes, at which requests are picked up and dropped off. As stops for pick-up and drop-off should not hinder traffic flow on major roads, only nodes without direct connection to roads, which have speed limit greater or equal

to 60 km/h, serve as boarding points.

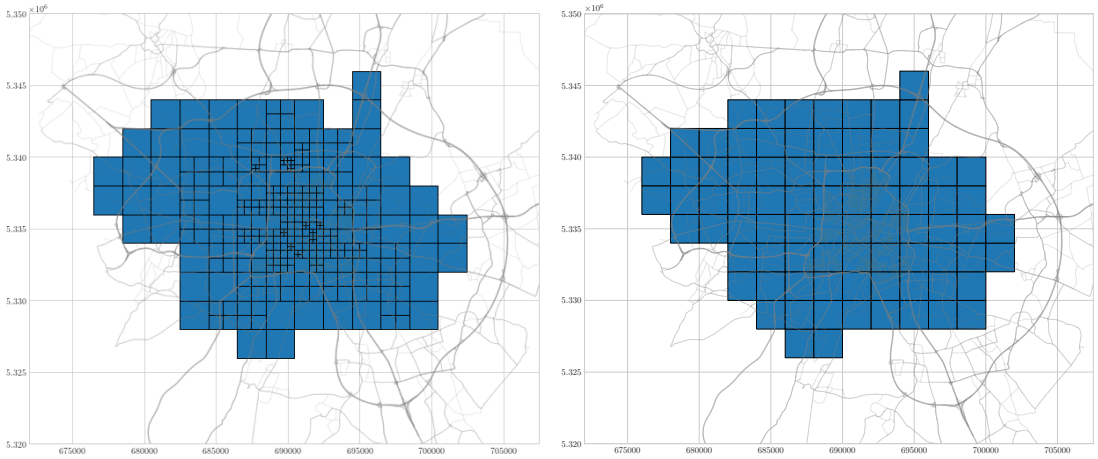
The AMoD demand is derived from OD matrices originating from a microscopic traffic simulation model [DANDL, BRACHER, et al., 2017] for private vehicle trip demand and a macroscopic traffic model of Bavaria [MAGET et al., 2019] for public transport demand. For the operational part of the thesis, demand is exogenous and assumed to be 5% of the combined private vehicle and public transport OD matrices. The number of requests sums up to more than 115 thousand trips and has peaks in the morning and late afternoon hours and is nearly zero during between 01:00 and 03:00. To avoid ignoring trips starting shortly before midnight, the simulation time is chosen to be from 03:00 of one day to 03:00 of the next. The time-based simulation uses a time step of $\Delta t = 1$ s. Request files are generated by applying Poisson processes on the hourly OD data to create explicit requests that enter the simulation as exogenous input at access nodes at the respective time steps. For simplicity and a clear evaluation, it is assumed that each request travels alone, i.e., $\rho_r = 1$, and that the parameters $\tau^w = 5$ min and $\tilde{\tau}_r^\delta = 40$ % are homogeneous[13]. For scenarios with reservation, additional request files are generated, in which $25, 50, 75$ % of requests are reservation requests with $\tau_r^e = 30$ min. The maximal fare f_r^{max} a request r is willing to pay for the AMoD service is created by computing the standard fare $f_r = \min(f^B, f^D \cdot d_r)$ and drawing a maximal acceptable dynamic pricing factor such that the demand curve reflects the logistic function in equation (3.59) with the choices $a = b = 1$. Additionally, the operator-side dynamic pricing threshold $\tilde{\beta}^p$ is set to 2. For the computaton of operator profits, explicit fare and cost values have to be set for this case study. The base fare f^B and the distance-dependent fare f^D are assumed to be 1 € and 1 € /km, respectively. The revenues based on these fare parameters face costs consisting of fix vehicle costs $c^F = 40$ € per day and variable vehicle costs $c^D = 0.25$ € /km.

In this chapter, constant travel times reflecting free-flow conditions are used to compare different operating strategies. With this simplification, it is possible to draw the trip duration distribution before simulations, which is illustrated in Fig. 3.14b. Most trips last between 5 and 10 minutes and approximately 94 % of all trips are shorter than 15 minutes. Fig. 3.14c displays the number of requests per 15 minutes. For comparison, yellow taxis in Manhattan have a mean trip duration of approximately 15 minutes and the number of requests reaches up to 4000 per 15 minutes[14]. Therefore, the mid-term time step Δt^M is selected to be 15 minutes. In order to keep the amount of time horizons small, the short-term horizon T^S and the mid-term forecast step size $T^{M,1}$ are also chosen to be 15 minutes. The total mid-term horizon is 1 hour in most simulations.

Various zone-systems are defined. For most simulations, the zone system in Fig. 3.15a with maximum edge length of 4 km and 4 aggregation levels is used. It contains squared zones with different sizes, which are determined by the demand density originating and ending within the respective zones. Zones, in which more demand is expected, are chosen to be smaller in order to allow smaller distances between the centroid and any other node, and zones with lower demand are chosen to be higher to save computational resources and have forecast values in

[13] Both simulation environment and fleet control strategies are capable of working with heterogeneous values, but the evaluation is more tractable for the homogeneous case.

[14] Evaluation of data from week 46 of 2018 by NYC TLC trip records https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page



(a) Multi-level zone system with maximum edge length of 4 km consisting of 121 zones.

(b) Constant size zone system with edge length of 2 km consisting of 82 zones.

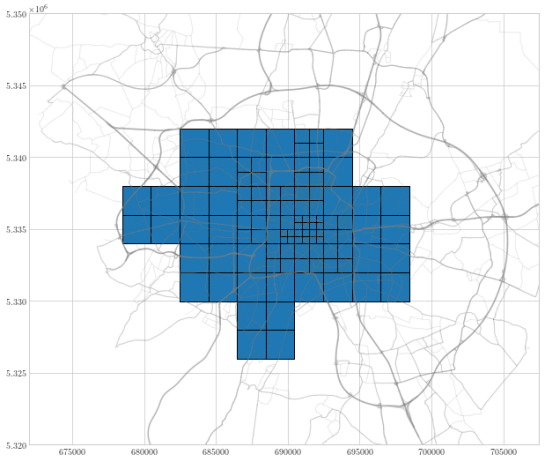
Figure 3.15: Zone-system definitions in this case study.

a similar scale. The algorithm for the creation of the multi-level zone system is described in more detail in Appendix B.4. This process generated 121 zones for a maximum edge length of 4 km. The same procedure was repeated for a maximum edge length of 2 km, resulting in 340 zones. Additionally, zone systems containing only squares (see Fig. 3.15b) with constant edge length were created for edge length of 4, 2, and 1 km, which created 26, 82, and 268 zones, respectively.

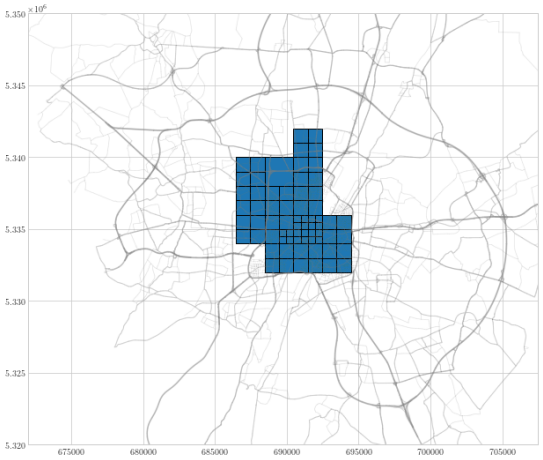
During the zone-creation process, forecast files with values for Λ_z^j and $\bar{\Lambda}_z^j$ with $j \in \{0, ..., 95\}$ are created by aggregating the trips originating and ending in the zones for every 15-minute interval j, respectively. Additionally, a second type of forecast file was generated by taking the mean of all demand files generated from different random seeds for each period and zone. The root mean squared deviation of the forecasts generated in this way is approximately 5.5 and the time evolution follows the trip rate curve (Figure 3.14c), i.e., errors are larger when there is more demand.

In order to pre-process the zone-to-zone reachability correlations $K_{z,z'}$, the required integrals in equations (3.33) and (3.35) are evaluated on a grid with 100 m distance between points and a cut-off at the boundaries of the outer zones. $K_{z,z'}$ are computed for reachability bandwidth parameters $h \in \{1, 500, 1000, 1500, 2000, 2500\}$ m. $K_{z,z'}$ is a unit matrix for $h = 1$ m; for the 4-lvl zone system with 4 km maximal edge length, the share of the non-diagonal elements from the total sum of all matrix elements is approximately 0, 25, 45, 58, 67, and 74 % for increasing values of the bandwidth. For other zone systems with larger (smaller) zones, the values are smaller (larger).

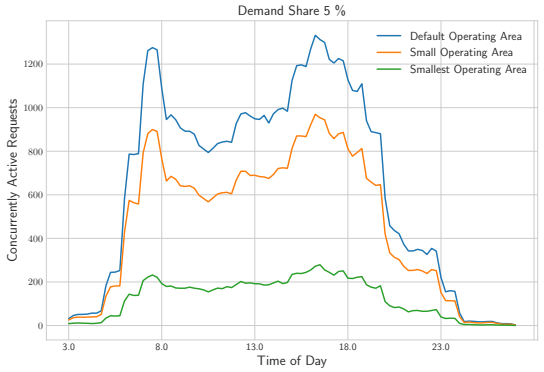
For each zone, the boarding point closest to the geometrical center serves as centroid. In most simulations, it is assumed that AMoD vehicles can find a parking space where they become idle; this can be achieved by reserving a few parking lots on most streets. Another approach is tested in a few scenarios, in which the operator has to return idle vehicles to depots, which are assumed in the centroids of the zones. In these scenarios, the vehicles also



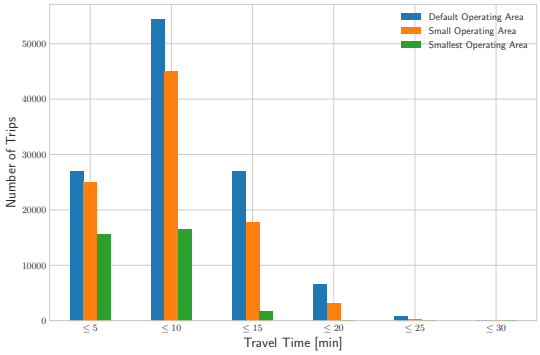
(a) Operating area without level 0 (lowest demand density) zones.



(b) Operating area without level 0 & 1 (lowest demand density) zones.



(c) Number of concurrent travelers in the three operating areas.



(d) Travel time distribution for the three operating areas.

Figure 3.16: Reduced operating areas and attributes of demand therein.

start in these depots; in all other scenarios the vehicles are distributed at the beginning of the simulation according to an initial distribution file derived from the trip origins between 00:00 and 05:00.

The question, where an AMoD service can be offered profitably is very interesting for many stakeholders. In a free market, operators are likely to use a geofence to optimize their profit. Additionally, cities require an understanding where they want the mobility service to operate and AMoD service providers should know if they have to ask for subsidies if part of that area is likely not profitable. In this thesis, three different operating areas are tested in this case study. For simplicity, these operating areas are generated from the available multi-level zone system shown in Fig. 3.15a, which classifies the zones by total demand density. Therefore, removing the largest, lowest-level zones is a simple method to build an operating area with higher trip density. The resulting operating areas and zone systems from performing this step twice are illustrated in Fig. 3.16a and Fig. 3.16b. In the second step, one level-1 zone was

not removed to keep a rather compact operating area[15]. The sizes of the respective operating areas are 310, 208, and 56 km^2. For the largest operating area, the 310 km^2 relate to the area of the city of Munich rather than the total area of the forecast zones. As mentioned in the description of equation (3.63), the number of concurrently active requests $|V|_<$ is a lower threshold for the number of vehicles that are required in a hailing operation. The quite large reductions in concurrently active requests for the different operating areas shown in Fig. 3.16c has two origins: first, the number of trips is reduced by both trips that originate and end in the removed zones; second, due to the smaller operating area, longer trips are removed from the trip duration distribution (as observable in Fig. 3.16d). By design, the request density is higher for smaller operating areas. The respective densities are 15.6, 18.3, and 25.2 requests / km^2.

3.6.3 Scenario Setup

The lowest fleet size for the hailing scenarios with 5 % penetration rate as demand is chosen such that the lowest value approximately matches the lower-bound estimation $|V|_<$. From there fleet size is incremented by 250 vehicles to analyze the sensitivity to fleet size. For the pooling scenarios, a smaller fleet can suffice depending on the average occupancy. Therefore, coarser vehicle increments of 500 vehicles are selected. The fleet is varied between 500 vehicles and 1500 vehicles, which should approximately be required for average occupancy rates of 1 to 3 passengers.

There are too many strategies and input parameters to allow for a global sensitivity analysis, in which the sensitivities of all strategies would be tested for different values of the input parameters. Hence, specific scenarios, in which a subset of strategies and input parameters are varied to analyze their specific impacts, are set up and described in the next section.

The first subsection focuses the analysis on different assignment algorithms without repositioning. Then, the reservation method is studied. Third, different repositioning strategies are compared. The impact of one repositioning strategy on the different assignment algorithm results is investigated before the sensitivity of the AMoD system performance to several hyperparameters is tested for a single repositioning strategy. Different zone systems and parking strategies are studied in the sixth subsection. The dynamic pricing strategies are included in the seventh section before finally, the results are analyzed for different operating areas.

For a quick overview, Table 3.3 presents a summary of default scenario parameters and in which of the subsections of section 3.7 they are varied.

[15]The result smallest operating area is actually very similar to today's carsharing operating area.

Parameter/Strategy	Variation	Default Value / Strategy
number of random seeds	-	3
time step	-	$\Delta t = 1$ s
simulation time	-	03:00 - 03:00 (86,400 s)
travel times	-	constant (free-flow)
street parking	6	allowed
initial vehicle distribution	6,8	constant (distributed in operating area)
demand	1,2	5 % of total private vehicle (PV) and PT demand
max. waiting time	1	$\tau_r^w = 5$ min $\forall r$
max. detour factor	-	40 %
reservation share	2	0 %
reservation time	-	30 min
boarding time	-	30 s
minimum fare	-	$f^B = 1$ €
distance fare	-	$f^D = 1$ € /km (computed on cent level)
fix vehicle costs	1	$c^F = 40$ €
distance costs	-	$c^D = 0.25$ € /km
hailing fleet size	1	$\|V\| \in [1250, 1375, 1500, 1625, 1750]$
pooling fleet size	1	$\|V\| \in [500, 750, 1000, 1250, 1500]$
operating area	8	as shown in Fig. 3.14a
short-term horizon	2	$T^S = 15$ min
re-optimization frequency	-	$\Delta t^S = 30$ s
hailing assignment strategy	1,4	with re-optimization
pooling assignment strategy	1,4	without re-optimization
RV heuristics	1	$N^{RV,h} = 20$; $N^{RV,wl} = 5$, $N^{RV,al} = 15$
mid-term horizon	5	$T^M = 60$ min ($J = 4$)
mid-term strategy frequency	-	$\Delta t = 15$ min
zone system	6	as shown in Fig. 3.15a
forecast method	3,7	forecasts based on average of random-seed trip data
zone correlation bandwidth	5	1500 m
repositioning strategy	3	QDLSTS with $N^{esp} = 10$
dynamic pricing	7	-

Table 3.3: Default parameter values and strategies for evaluated scenarios. The 'Variation' column shows in which of the subsections this parameter or strategy is varied.

3.7 Case Study: Results

3.7.1 Comparison of Assignment Algorithms

First, the impact of the global re-optimization and the scaling properties of the assignment algorithms is tested for both hailing and pooling when repositioning is not active. Table 3.4 summarizes the scenario design. All scenarios utilize an immediate response system (IRS) to

Category	Description	Abbreviation / Variation
assignment strategies	hailing without re-assignments	HailIRS
	hailing with re-assignments	HailIRSBatch
	pooling without re-assignments	PoolIRS
	pooling with re-assignments	PoolIRSBatch
RV heuristics	with / without RV	
demand level [%]	share of total PV and PT demand	1,5,10,15
fleet size per % demand	hailing scenarios	[250,275,300,325,350]
	pooling scenarios	[100,150,200,250,300]

Table 3.4: Scenario parameter variation for comparison of assignment algorithms.

generate offers and make initial heuristic assignments in case the request accepts. In two thirds of scenarios, periodic re-optimization of request-vehicle assignments is performed in batches every $\Delta t^S = 30$ seconds, which is denoted by *IRSBatch*, in the other third, vehicle plans are only changed by the heuristics that generate the offers, which is denoted by *IRS*. Moreover, RV-heuristics to speed up computations are applied in half of the scenarios with periodic re-optimization: in the hailing case, the number of RV connections is limited to $N^{RV,h} = 20$ and in the pooling case, $N^{RV,wl} = 5$ and $N^{RV,al} = 15$ are chosen. To test the scaling properties, both demand and fleet size are varied: additional to the 5 % demand scenario, 1, 10, and 15 % of the total private vehicle and public transport traffic are treated as requests to the AMoD system. The scenarios are simulated for fleet sizes of 250, 275, 300, 325, and 350 vehicles per % of demand.

The computation time is an indicator for the complexity of the solution approaches. For real-time applicability, the operator processes have to be performed reasonable fast, i.e., optimization processes have to finish before the next optimization is called. The total computation time is the sum of the time spent on IRS and batch re-optimization for a single simulation. For real-time applicability, the total computation time has to be lower than 24 hours for 24 hours of simulated time[16]. Parallelization of processes would allow improvements in computation time; however, as computational resources are limited for this thesis, the focus was to simulate multiple scenarios on a single core rather than focus on code efficiency and parallelization.

Fig. 3.17 gives a good impression of computation times. As expected, all simulation times increase with fleet size and demand level. A nearest-neighbor IRS policy for hailing is computationally very efficient and even 500 requests per minute (15% demand scenario) and fleet sizes of 5000 vehicles are computationally feasible. The other strategies besides this hailing IRS strategy, which has 3 scatter points per fleet size and demand level, were computed with and without RV heuristics and therefore have 6 scatter points in general. If less data points are visible, there can be two explanations: (i) the data points are indistinguishable or (ii) the simulations lasted more than a maximum run time of 40 hours, in which case no data points

[16]In fact, the computation time has to be even lower, as the computation during peak hours lasts longer than the computation in off-peak/night hours.

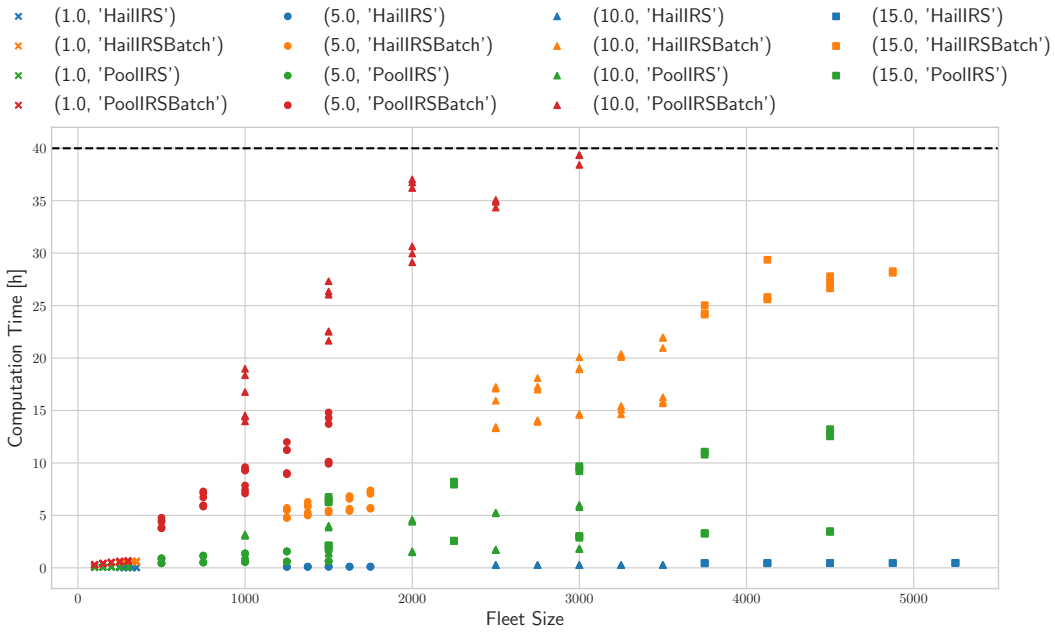


Figure 3.17: Scatter plot for total computation time of various assignment algorithms for all random seed simulations with different fleet sizes (x-axis). The legend entries refer to demand-levels (markers) and assignment strategies (colors).

were created. In general, the scattering for simulations with different random seeds is very small and points are often hardly distinguishable. The next lowest computation times were achieved by the pooling IRS strategy. This strategy is real-time capable on a single core even for the 15 % demand scenario. The computational efficiency of the applied RV heuristics becomes obvious as with RV-heuristics, the computation time can be kept below 5 hours for 4500 vehicles, whereas the computation time is approximately 13 hours without RV heuristics. For the hailing case with batch re-optimization, the application of heuristics becomes the more important the larger the scale of the problem: without heuristics, the 15 % demand scenario does not finish within 40 hours of computation time. The significant scaling of ride-pooling possibilities with demand becomes apparent when looking at the simulations with ride-pooling re-optimizations (PoolIRSBatch), e.g. by comparing the computation times of 5 and 10 % demand scenarios for 1000 or 1500 vehicles. Simulations with 10 % demand and fleet sizes of 2500 and 3000 vehicles did not finish in time without RV heuristics. For the 15 % scenario, not even the application of RV heuristics could limit the solution space enough to achieve a computation time below 40 hours.

Next, the profit objective is evaluated. Fig. 3.18 shows a scatter plot for the simulated scenarios. As expected, profit increases with demand level and pooling generates higher profits than hailing. The differences between different fleet sizes for the same demand are rather small, but profit increases with fleet size for the studied fleet sizes. However, the curves flatten for higher fleet sizes, indicating that the optimal fleet size without repositioning is nearly reached[17].

Moreover, the blue and the orange symbols and some red and green symbols overlap.

[17]See Fig. 3.25 for scenarios with larger fleet sizes for the 5 % scenario.

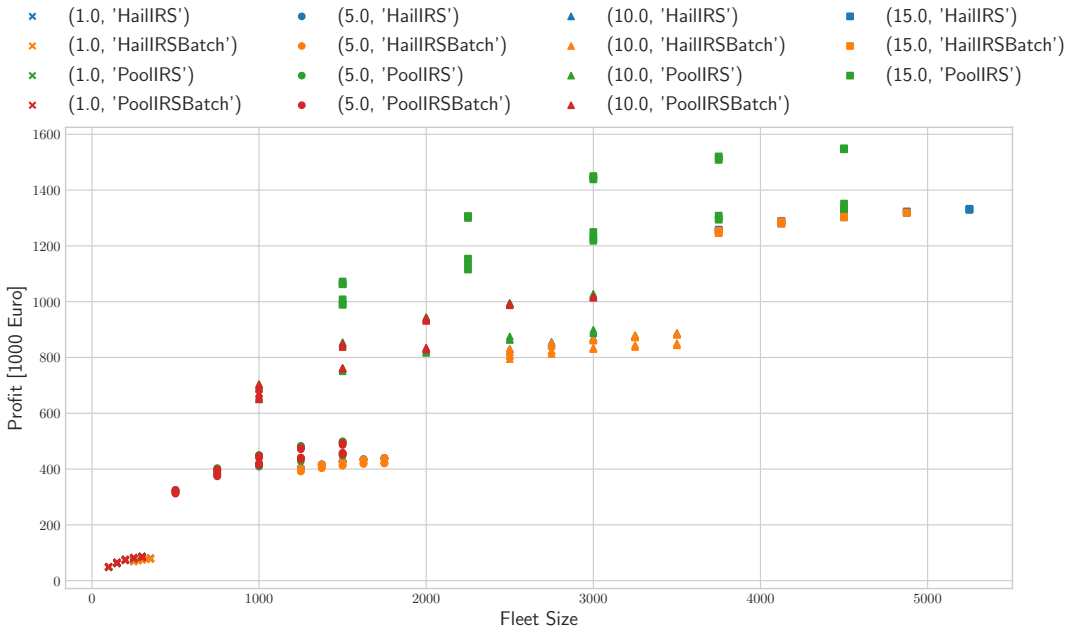
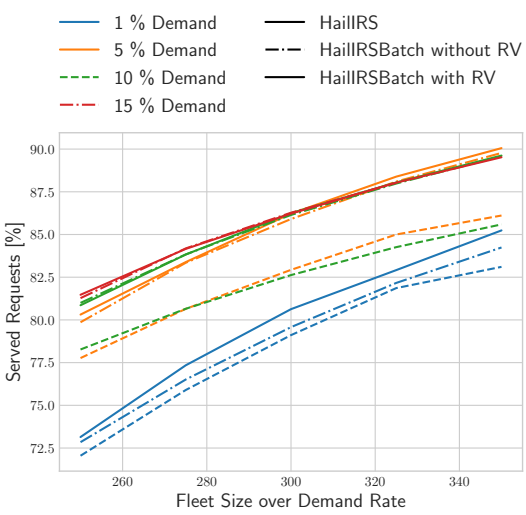


Figure 3.18: Scatter plot for operator profit of various assignment algorithms for all random seed simulations with different fleet sizes (x-axis). The legend entries refer to demand-levels (markers) and assignment strategies (colors).

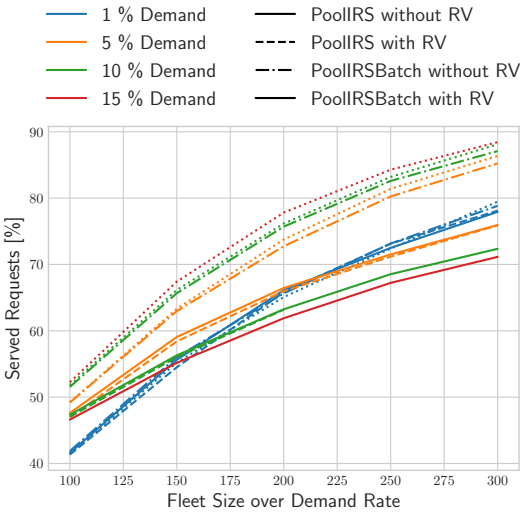
Therefore, it can be observed that the profits of simulations with and without re-optimization are very similar. Contrarily, the RV heuristics cause quite large deviations, especially for pooling and larger scales (demand and fleet size).

The evaluation of both the computation time and profit shows the stability of results over demand data sets from different random seeds. In most cases, the three scatter points for the three random seeds are hardly distinguishable. This is also true for other KPIs, which is why for the rest of the evaluation, mean values of the three random seed simulations are discussed for a clearer description. An explanation for this result could be that overall, the trip patterns in the request files have a very high similarity as they originate from the same OD matrices and a very large number of Poisson processes, namely more than $100,000$ per simulated day for the 5 % scenario.

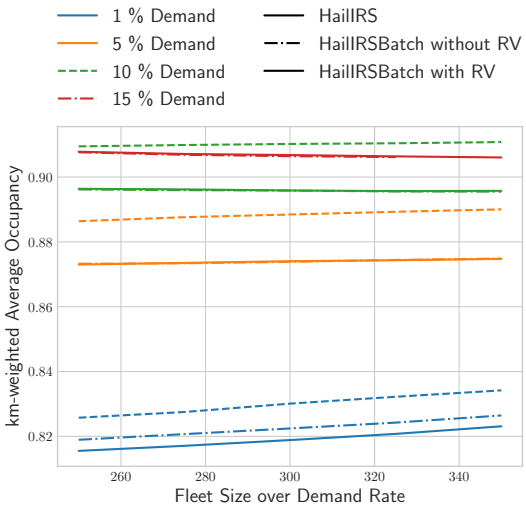
In the scenarios studied in this section, the profit for the strategies without re-optimization or RV-heuristics is very similar to those with more complex operator strategies. This can be traced back to the dynamics of the original problem; future demand in the dynamic problem might be better served by a solution that is not optimal according to the current state without forecast information. Indeed, observing the different line types in Fig. 3.19a and Fig. 3.19b shows that more requests are served in scenarios with RV heuristics. Note that the HailIRS strategy inherently only considers the vehicle plan for a single vehicle. These two diagrams also illustrate the scaling behavior with demand and fleet size. For all demand levels, the curve of served requests does not yet show saturation over fleet size. The macroscopic fleet size estimation was too optimistic for strategies without forecasts: approximately 80 % of demand can be served for the best hailing and pooling scenarios with 250 vehicles per % of



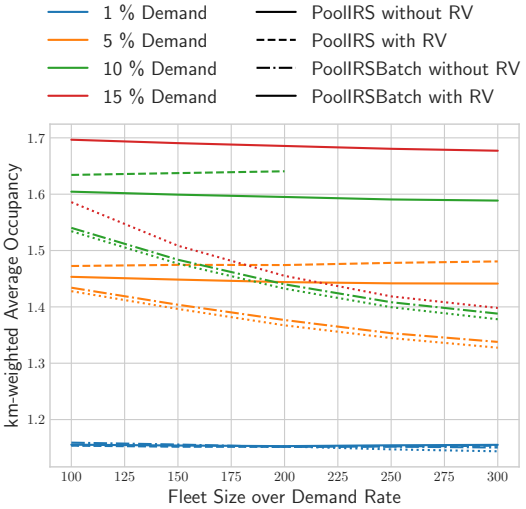
(a) Share of served requests for hailing scenarios.



(b) Share of served requests for pooling scenarios.



(c) km-weighted average occupancy for hailing scenarios.



(d) km-weighted average occupancy for pooling scenarios.

Figure 3.19: Shared of served requests and average occupancy for variations of demand, fleet size, and assignment strategy.

demand. However, fleet sizes of 350 vehicles per % of demand are still not sufficient; this is a consequence of demand patterns and vehicle demand and supply imbalances, as vehicles might be available but not within the maximum waiting time radius of requests that cannot be served. Fig. 3.20 displays the vehicles that remained idle for 15 minutes on the one hand, and the location of unserved requests during that time period on the other hand, for a scenario with 5 % demand, 1750 vehicles, and the HailIRS strategy. The effect of availability in the wrong locations also becomes apparent from the evaluation of the peak fleet utilization, which is roughly 81 % in this scenario, i.e., 20 % of vehicles are not used while users cannot be

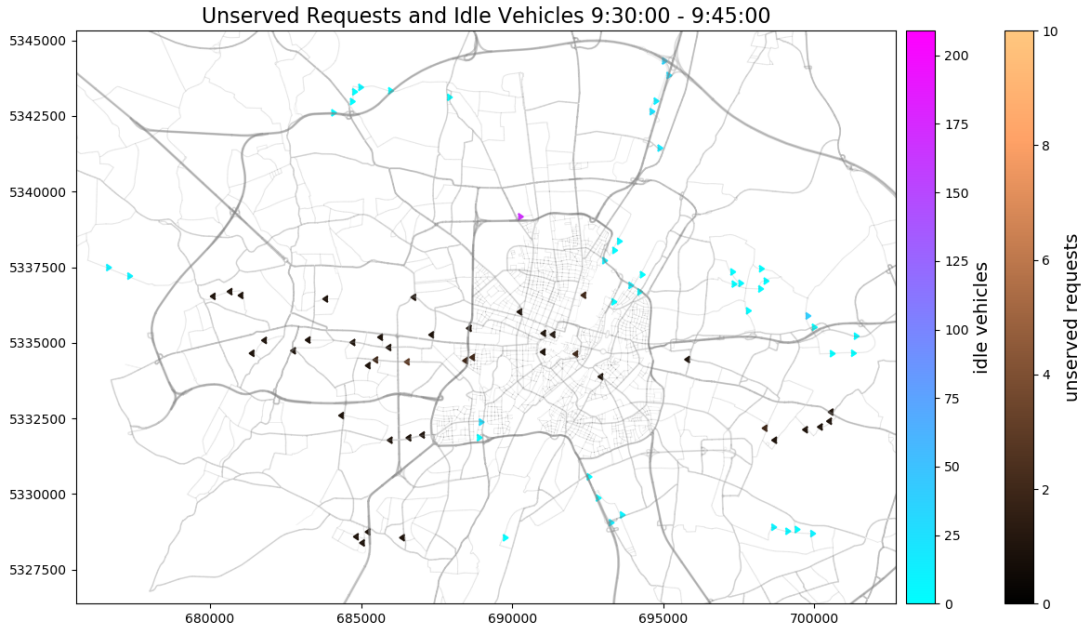


Figure 3.20: On the spatio-temporal imbalance of demand and supply for the 5 % demand scenario with 1750 vehicles and the HailIRS assignment strategy.

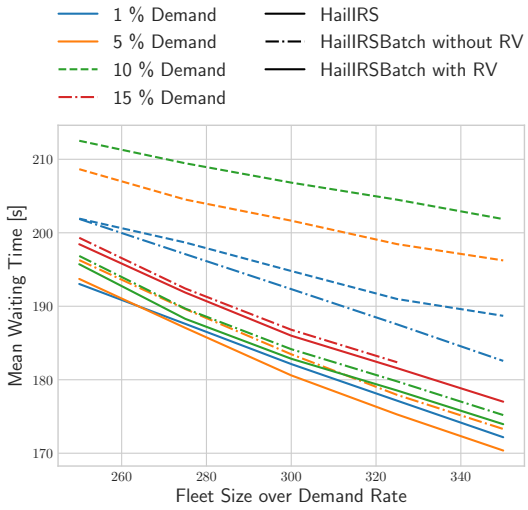
served. The scaling behavior of served requests with demand and fleet size is very similar for hailing scenarios with more than 5 % demand; only in the 1 % demand scenario, the operator is not able to serve the same share of requests. This result actually can be expected as the probability of finding an available vehicle nearby is higher for scenarios with a higher vehicle density. Obviously, the probability of finding any available vehicle within a certain radius is also a function of the ratio of fleet vehicles to requests. The km-weighted average occupancy in the hailing case (see Fig. 3.19c) is determined by the empty pick-up trips and reflects the distance to the next available vehicle. In the 1 % scenario, an increase in average occupancy and a decrease in average pick-up distance is observable. Interestingly, this effect is not observable in the scenarios with more demand, where the results are stable for different fleet sizes. For the shown fleet sizes, the system is still in an under-supply state. As more vehicles are put into the system, the pick-up trip distance is mostly determined by the demand density. Therefore, it is also clear that the km-weighted average occupancy is increasing with demand level in these scenarios. For hailing systems, a trade-off becomes apparent: re-optimization without RV heuristics finds solutions with shorter pick-ups; however, without any repositioning, the non-myopic nature of this optimization brings the system into states that are not favorable for the future.

The results for pooling are shown in Fig. 3.19b and 3.19d. In the computed scenarios, the approaches with and without re-optimization can serve approximately the same amount of requests and the re-optimization can do so with a higher km-weighted average occupancy. The application of RV heuristics, which limits the number of considered vehicles to only 20 in this case study, severely limits the solution space for assignments. The effect is similar to the hailing case: the heuristic approach actually can serve more requests but at decreased pooling efficiency. The decreasing average occupancy with fleet size can be explained by two effects.

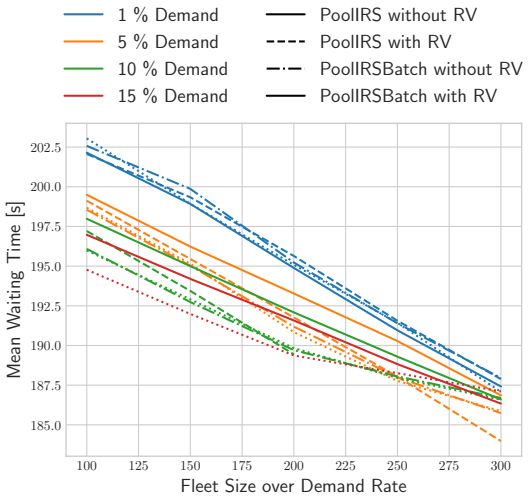
First, the constant number $N^{RV,al}$ as with increasing fleet size the number of available vehicles increases and the probability of getting the vehicle with the best match decreases. Second, the requests that are served additionally with larger fleets, might be requests that are not matching well with other requests. This likely is also the reason for a small decrease in average occupancy for the IRS strategies without RV heuristics. The re-optimizations without RV heuristics even have a slight trend with fleet size even though more users are served, highlighting the large optimization potential for ride-pooling assignments. The scaling behavior with demand shows the same trend for a pooling system as for the hailing system, however, the efficiency measured by average occupancy scales much stronger. Comparing the strategies without RV heuristics, the relative difference in average occupancy between the 1 and 10 % demand scenario is within 10 % for hailing, whereas this difference is approximately 30 % for pooling.

As observable in Fig. 3.21c, the evaluation of saved distance shares its trends with the km-weighted average occupancy. The value of saved distance is approximately 32 % for the 15 % demand scenario with the PoolIRS strategy and RV heuristics. If there would be no detours at all, the average occupancy of 1.7 in the 15 % demand scenario could translate to a saved distance of approximately $(1 - 1/1.7) \approx 41$ %. The results show a saved distance of roughly 32 % indicating very reasonable detours. Since the operator objectives are to minimize unserved customers and driven distance and the users have detour constraints, unnecessary detouring with multiple users on board is avoided. The detour time, which contains time delays due to detours and boarding processes, also follows these general trends. Scenarios with RV heuristics have shorter detours than their counterparts without RV heuristics and the detours are decreasing with increasing fleet size as the probability for pooling is decreased (see Fig. 3.21d). In general, scenarios with more demand have longer detour times: the mean relative detour time is $17 - 19$, $22 - 24$, $24 - 25$, and approximately 25 % in the IRSPool scenarios with 1, 5, 10, and 15 % demand, respectively. The share of empty VKT is exactly correlated with the km-weighted average occupancy in the hailing case and also shows the same trends in the pooling case. Due to the possibility to insert new request pick-ups at any point of the route, the share of empty VKT is lower for pooling without RV heuristics (4 to 13 %) than for hailing ($9 - 18$ %) for the same level of demand. Applying the selected RV heuristics for pooling increases the share of empty VKT considerably, especially for larger fleet sizes (9 to 15 %). Fig. 3.21a and 3.21b show that the mean waiting time is rather stable if demand and fleet size are scaled the same way. The differences between assignment strategies are also rather small. For 300 vehicles per % of demand, the mean waiting times are between 180 and 190 seconds in almost all scenarios; longer mean waiting times of up to 210 seconds only occur for the IRSBatch strategies in a hailing operation. In general, the waiting times decrease with the ratio of fleet size over demand.

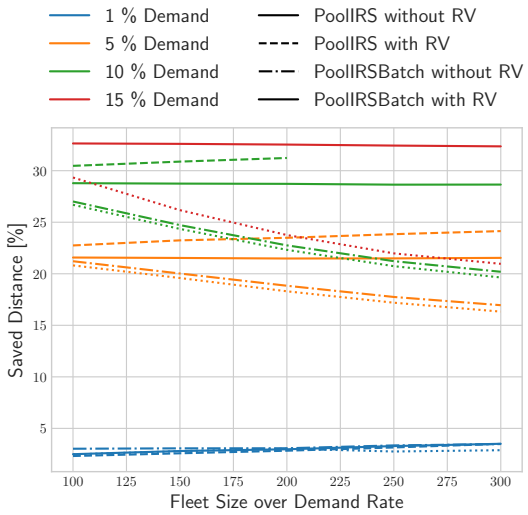
Fig. 3.22 illustrates the occupancy over the course of a day for four scenarios with the same fleet-size-to-demand ratio. This highly aggregated stackplot is very useful as it shows both the total fleet utilization as well as the distribution of trips with respect to occupancy over time. In comparison to the 1 % demand scenario with 300 vehicles, the 10 % scenario with 3000 vehicles has a smaller share of empty trips and additionally, the shares of trips with 2, 3, or 4 passengers are higher. The trend of more demand leading to better pooling rates is also apparent during a single simulation as the share of pooled trips is higher during peak hours. The 5 % demand case shows a higher similarity to the 10 % case than the



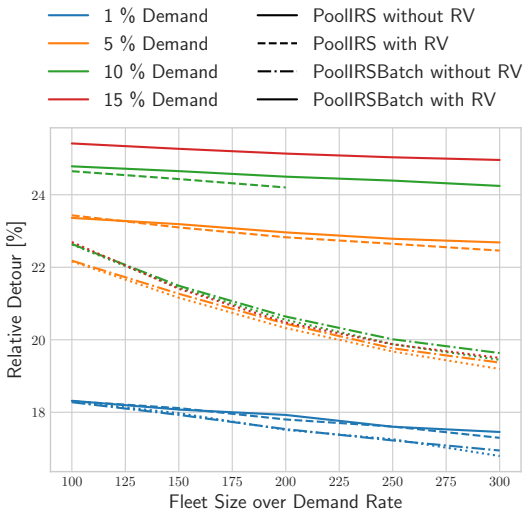
(a) Mean customer waiting time for hailing scenarios.



(b) Mean customer waiting time for pooling scenarios.



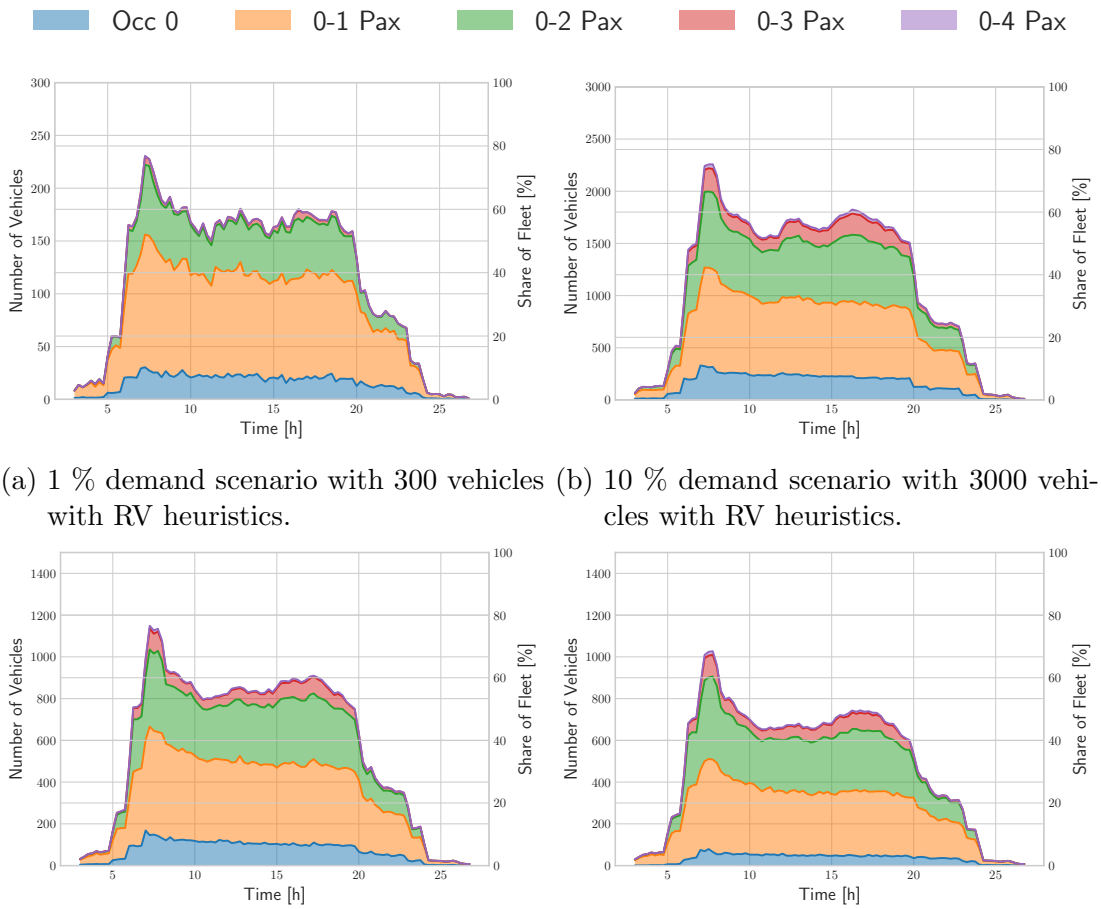
(c) Relative saved distance for pooling scenarios.



(d) Mean customer relative detour time for pooling scenarios.

Figure 3.21: Evaluation of further KPIs for variations of demand, fleet size, and assignment strategy.

1 % case. In all four scenarios, the fleet is utilized less than 80 % during peak hours even though there is unserved demand. In other words, more than 20 % of vehicles are not available where they would be needed to serve additional requests. This becomes especially apparent in the afternoon hours, where the fleet utilization is around 60 % for the scenarios with RV heuristics. It is not unexpected that this happens for all demand levels as for the same assignment strategy the vehicles follow the same demand patterns, i.e., the spatio-temporal distributions of requests' origins and destinations leading to the imbalance of demand and supply, in strategies without repositioning. A comparison of Fig. 3.22c and 3.22d suggests



(a) 1 % demand scenario with 300 vehicles with RV heuristics.

(b) 10 % demand scenario with 3000 vehicles with RV heuristics.

(c) 5 % demand scenario with 1500 vehicles with RV heuristics.

(d) 5 % demand scenario with 1500 vehicles without RV heuristics.

Figure 3.22: Stack plots showing vehicles with certain occupancy for IRS pooling scenarios. The 0-occupancy vehicles are divided into empty trips (blue) and idle vehicles, which are represented by the white area on top.

that dynamic effects of selecting different request-vehicle assignments worsen the situation in the scenarios without RV heuristics. Even though the strategy without RV heuristics is more efficient in the morning peak by serving a similar share of requests with less vehicles, the resulting distribution of vehicles is disadvantageous for the afternoon and evening hours. If demand could be served, the vehicle utilization or the pooling rate would reflect the afternoon surge visible in the graph for concurrent travelers in Fig. 3.16c. Instead, more vehicles are in places where they cannot be assigned to new requests thereby increasing the share of idle vehicles and decreasing the share of requests that can be served. The plots for the pooling strategies with re-optimization (IRSBatch) are very similar as for the IRS scenarios, which is why they are not shown; the RV heuristics in the selection of accepted requests determines which requests are served and the optimization potential afterwards seems rather limited.

After this global sensitivity analysis for assignment algorithms, some more restricted parameter variations were set up to test

- how a cut-off at a larger maximum waiting time would change the computational effort, the operator profit, the number of unserved requests, the mean waiting and mean journey times

- how much more demand could be served with larger fleet sizes and under which circumstances operators would consider this by approximating the sensitivity to variable and fix vehicle costs

For a variation of maximum waiting time, demand and fleet size were set constant to 5 % and 1500 (hailing) / 1000 (pooling), respectively. The computation times for maximum waiting time values of 5, 7, and 10 minutes are shown in Fig. 3.23a. Increasing the maximum user waiting time τ^w clearly increases the computational effort as more vehicles can reach a new request r within this time and can be considered for assignment. The scenario with $\tau^w = 10$ minutes cannot even be computed within 40 hours with pooling and re-optimization on a single core. RV heuristics limit the negative scaling effect of increased waiting times on the computation time. As these heuristics act before the different stop insertions are checked for feasibility and control objective value, their effect is especially severe for pooling.

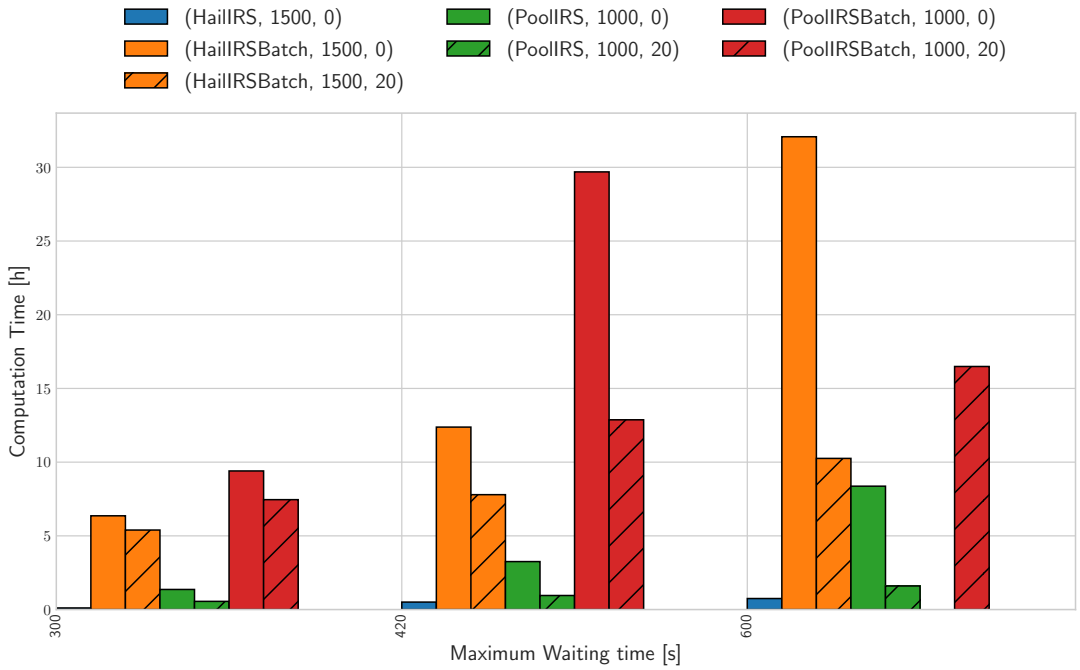
Expectantly, the average waiting times increase with increasing maximum waiting time (see Fig. 3.23b). It can be observed that the hailing IRS approach and the strategies with RV heuristics utilize the additional possibilities less extensively.

As illustrated in Fig. 3.24a, the increase in maximum waiting time leads to an increase of profit by up to 20 % for the hailing case as more customers can be served (see Fig. 3.24b). Without repositioning, the fleet cannot be utilized to its full potential for shorter waiting times. This can be imagined by drawing circles of availability around the vehicles on a map; the radius is the maximum waiting time multiplied by an average speed[18]. If new requests are within these circles, a match will be possible. Hence, the probability of a match increases with increasing maximum waiting time. The share of served requests rises from 83 to 91 and to 99 % for maximum waiting times of 5, 7, and 10 minutes, respectively[19]. Most of this reasoning is valid for the sensitivity analysis in the pooling case as well, albeit the scenarios with RV heuristics show a slightly different behavior for 10 min waiting time. The restriction of considered vehicles cannot increase the share of served requests anymore but instead even decreases the average occupancy and thereby profitability.

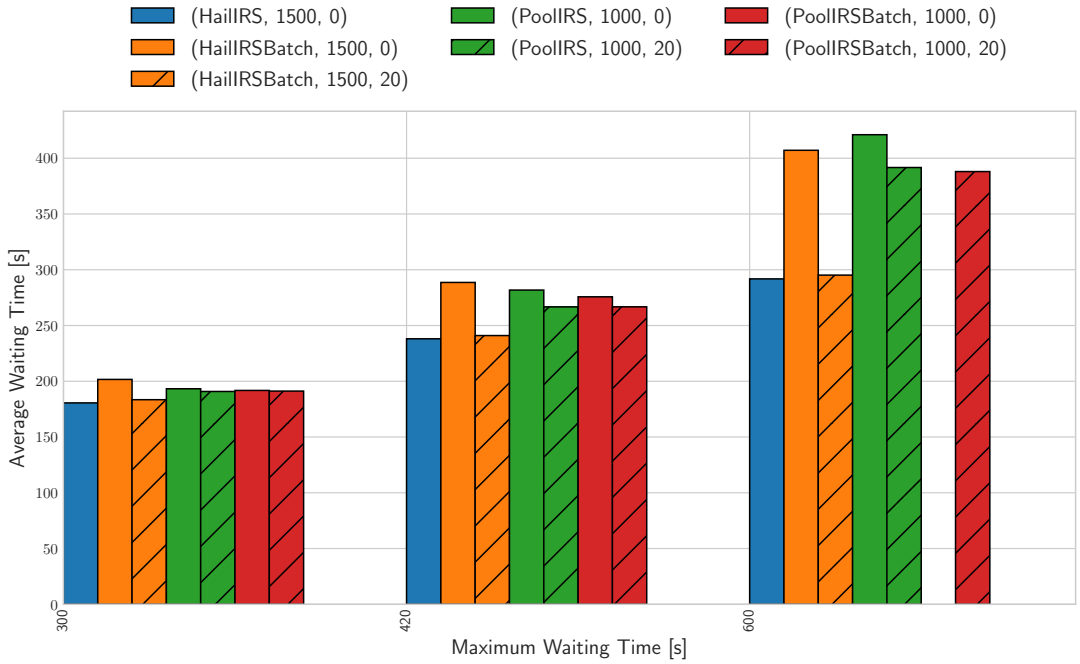
Another approach to serving more requests is to increase the fleet size (see Figure 3.25). To keep the computational effort manageable, simulations were only set up for the 5 % demand scenarios with RV heuristics. With increasing fleet size, the number of share of served customers grows expectedly. Interestingly, the served-request curves for hailing and pooling are nearly identical, with only slightly higher values for pooling. For 2500 vehicles, which is almost twice of the maximum value of the maximum number of concurrent travelers in the system (Fig. 3.16c), the share of served requests reaches only 95-96 %. Even more vehicles would be needed to counteract the demand-supply imbalance for short maximum waiting times. The sensitivity of fleet size on the mean waiting time is higher for hailing system. With more

[18] For en-route vehicles, the location is given by the next drop-off location and the radius is given by the maximum waiting time minus the time of availability.

[19] This validates the macroscopic approach to making first estimates for hailing fleet sizes given a demand profile.

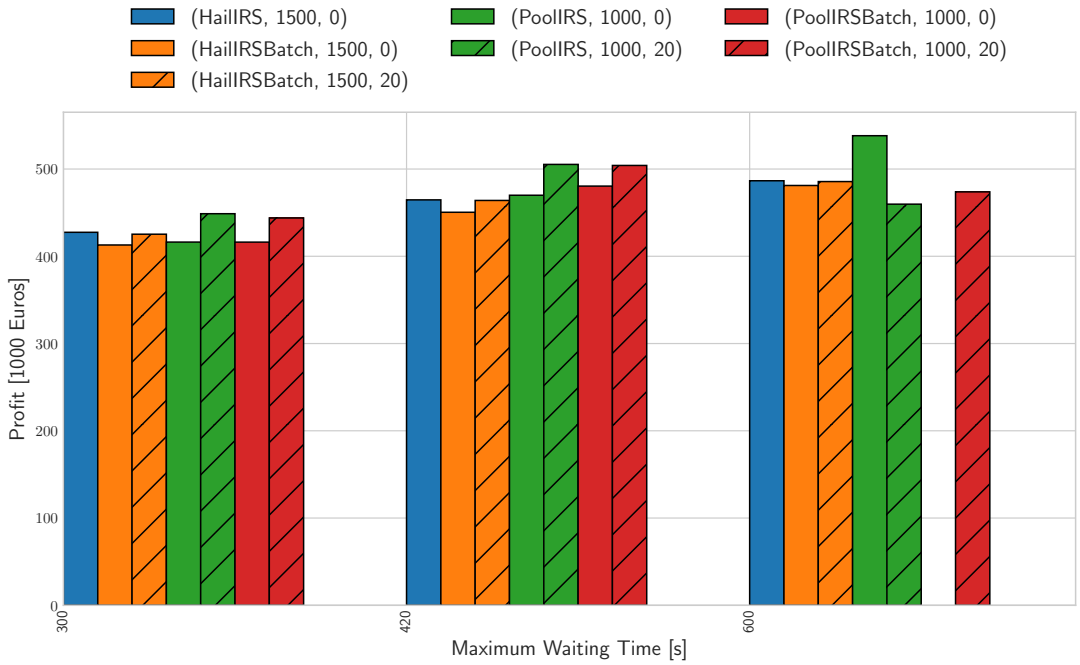


(a) Computation time.

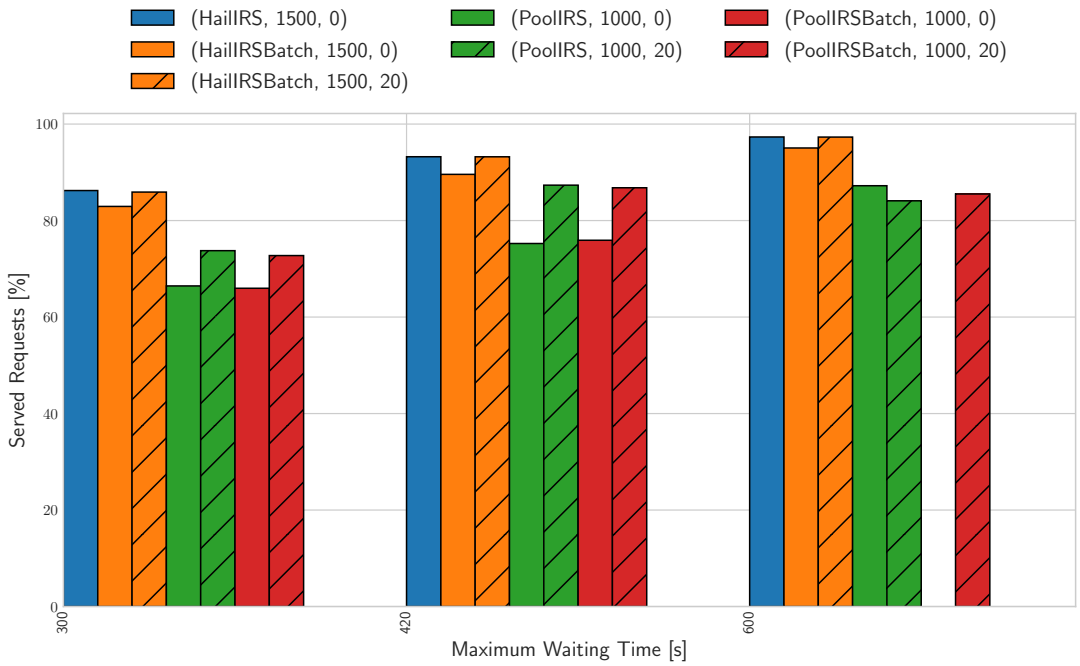


(b) Average waiting time.

Figure 3.23: Computation times and average waiting times for scenarios with different assignment strategies for a variation of maximum waiting times. The legend tuples refer to (assignment strategy, $|V|, \sum N^{rv}$).

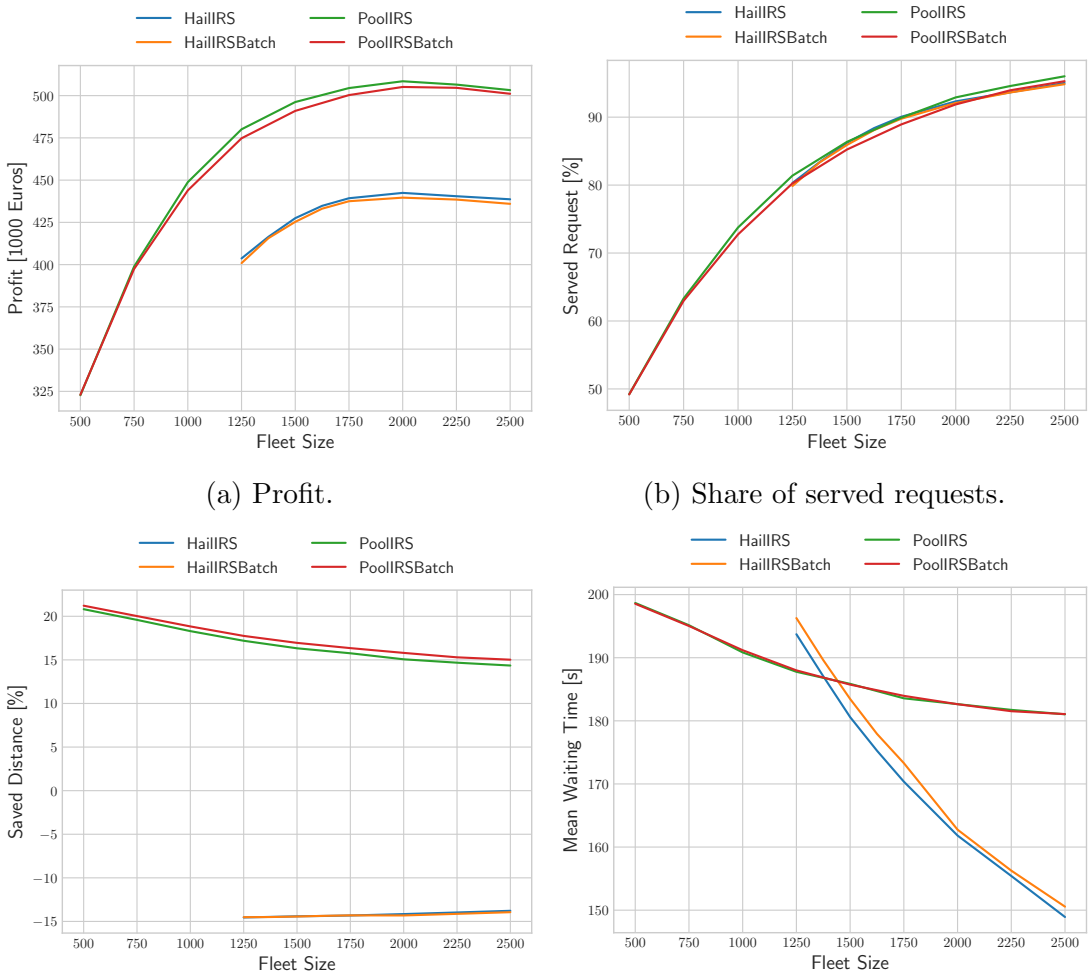


(a) Profit.



(b) Share of served requests.

Figure 3.24: Profit and share of served requests for scenarios with different assignment strategies for a variation of maximum waiting times. The legend tuples refer to (assignment strategy, $|V|, \sum N^{rv}$).



(a) Profit.

(b) Share of served requests.

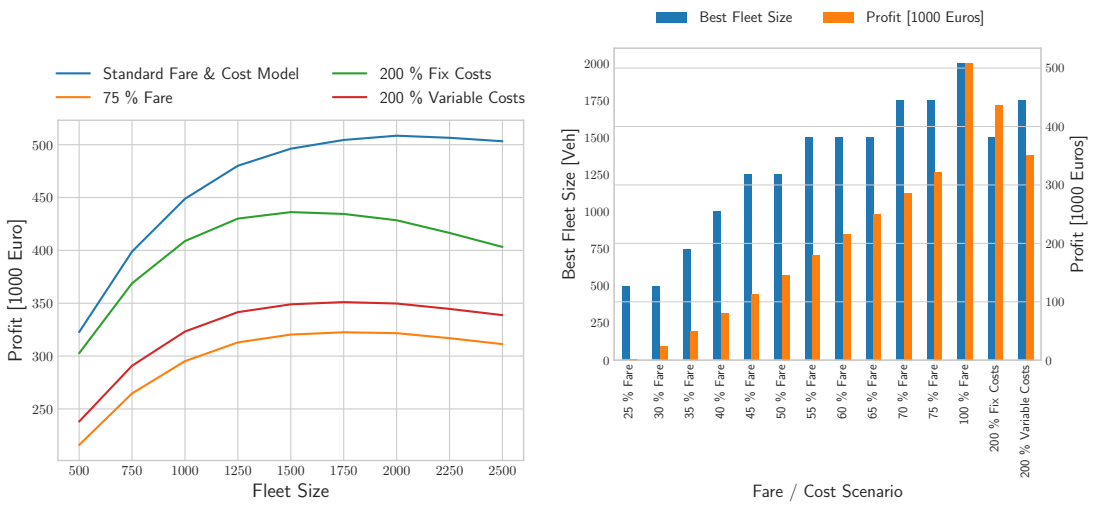
(c) Saved distance.

(d) Mean user waiting time.

Figure 3.25: Evaluation of larger fleet sizes for 5 % demand scenario with RV heuristics.

additional vehicles, the probability of finding a nearby available vehicle with short user waiting time increases, and the operator is likely to assign such additional vehicle due to the control objective. For pooling systems, the detours of users are also considered in the control objective function; hence, the probability of an assignment is depending less on the vehicle density and fleet size. Opposite trends can be observed for hailing and pooling with respect to the saved distance indicator. A higher vehicle density reduces the mean empty pick-up distance in hailing systems slightly whereas the pooling results show a negative impact of fleet size, which is not inherently a result of pooling but can be attributed to the RV selection heuristics (as described with Figure 3.19d). Nevertheless, this direct comparison shows the benefits of pooling over hailing: instead of generating approximately 15 % additional VKT, pooling can reduce VKT by 15-20 % (compared to the case where PKT = VKT). This is also in the interest of the operator. Serving the same amount of users with less VKT generates a surplus of profit as displayed in Fig. 3.25a.

Under the assumed cost and fare parameters, an operator would make large operational



(a) Variation of fleet size.

(b) Best fleet size and profit.

Figure 3.26: Variation of fare & cost parameters for 5 % demand scenario with IRSPool strategy with RV heuristics.

profits. However, there is a lot of uncertainty in the cost and fare parameters used for this case study. The variable, distance-dependent operating costs such as costs for energy and wear, could remain in a similar range as operating costs today. The fixed costs could be higher due to higher manufacturing costs for automation, at least in the introduction phase of this new technology. In order to see the impact of higher operating and fixed costs, each of them was increased by 100 % as a post-processing step, i.e., the simulations were not repeated with different objective functions, but the evaluation was adapted. Fig. 3.26a illustrates that the variable costs have more impact than the fixed costs. For the 5 % demand scenario with 1500 vehicles, a vehicle drove approximately 385 km to serve 66 requests on average per day. With this high mileage per day, additional investment costs of 10 € per day (e.g. for automation) translate to less than 0.03 € per kilometer of additional costs. An expected but noteworthy observation is that the increase in fixed costs results in a more convex fleet size to profit relation than an increase in variable costs. Moreover, the effect of reducing the fare or increasing the variable costs results in similar functional relations between fleet size and profit.

The fleet size with the highest operational profit and the respective profit are shown for given fare/cost systems in Fig. 3.26b. The fares are varied more extensively as they likely have higher uncertainty than the costs: the price elasticity of potential users is unknown, there could be market effects of competition or there might be regulatory measures in place, for example. Moreover, the operational profits will have to make up for overhead costs, e.g. for research, marketing, platform. Interestingly, the scenario with 25 % fare, for which the per-km fare and costs are both 0.25 € /km, still produces a slight profit due to small profits gained by a base fare of 0.25 € for trips shorter than one km. With increasing fare level level, the operator will put more vehicles into the system to serve more customers as the resulting surplus in trips makes up for the additional fixed costs.

3.7.2 Reservation Treatment

To check the necessity of a special treatment of reservation requests, the 1 % demand scenario is computed with a short-term horizon $T^S = 30$ min. With this setting, all reservation requests are considered in the re-optimization process until they are picked-up. However, the number of possibilities to assign requests to vehicles is very large as many vehicles reach a new request within 30 minutes. Additionally, requests that can be matched together create vehicle plans for many vehicles in the pooling case. As a consequence, the additional computational effort increases exponentially and not even the 1 % scenario with each second request being reserved 30 min in advance can be computed with re-optimization and a short-term time horizon considering full optimization of all known requests, i.e., larger than 30 min.

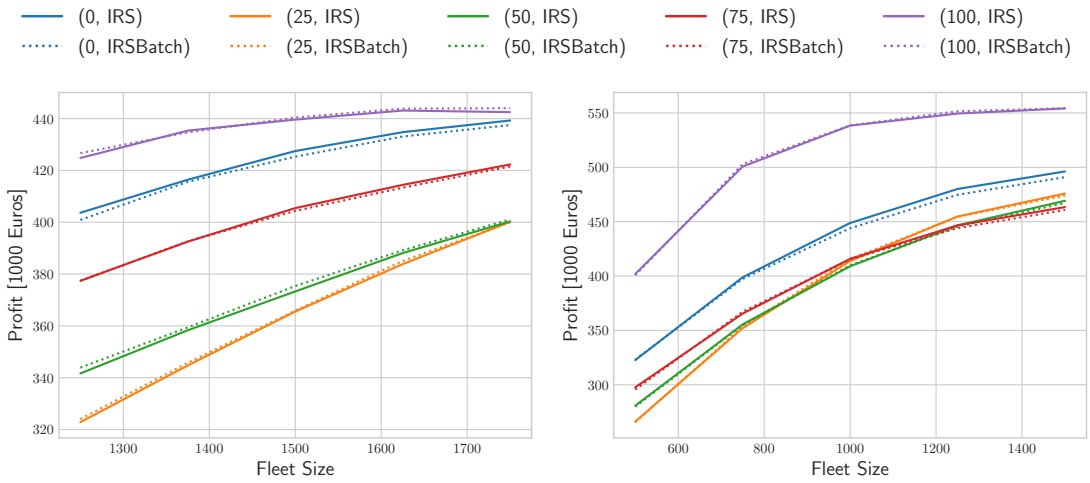
The methodology described in section 3.3.3, which only considers the requests starting within a shorter time horizon of $T^S = 15$ minutes in the global re-optimization step, is computationally feasible, at least for the tested 5 % scenario[20].

Fig. 3.27 illustrates results of scenarios, in which the share of reservation requests is varied. The results do not show a clear trend and require further interpretation. For both hailing and pooling, the 100 % reservation case shows the best results (from a profit point of view), followed by the 0 % case, and then the mixed reservation scenarios. The profit is directly related to the share of served requests in the respective scenarios. The pooling efficiency (Fig. 3.27f) measured by relative saved distance also has this order. Only the empty mileage in the hailing case, which is shown in Fig. 3.27e, deviates and the mixed scenarios are better in this KPI. Another finding is that the re-optimization potential seems to be rather limited in the studied scenarios with the chosen objective as all IRS and IRSBatch results are very similar. In the under-supply situation, the optimization is not capable of freeing resources in the regions, where vehicles (hailing) / routes with capacity (pooling) are needed, to accommodate new requests.

As expected, a comparison of hailing (Fig. 3.27a) and pooling (Fig. 3.27b) shows that a pooling system has a higher benefit of having only reservation requests over having only online requests. The additional foresight helps in building more efficient routes with more pooling, even for the heuristic IRS approach. This in turn allows a larger proportion of requests to be served. In contrast, the share of empty miles is higher in the 100 % hailing reservation scenario as the additionally served requests are reachable due to the larger search radius, which translates into longer pick-up trips.

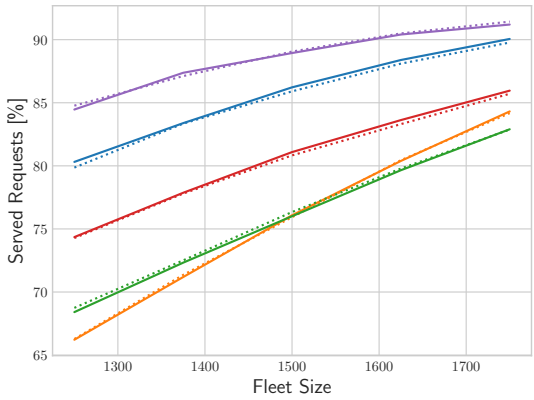
Compared to the online scenario, the share of served requests decreases by approximately $7 - 15$ % (25 and 50 % reservations) and $5 - 10$ % (75 % reservations) for the hailing case, where the difference becomes smaller as the fleet size increases. For the pooling case, online and partly reservation scenarios show similar levels of served requests, but the relative saved distance is better for the online scenario than for the scenarios with some online and some reservation requests. This can be traced back to the IRS being more likely to accept requests that fit into currently available vehicle plans in the online only case and avoid requests that would cause long empty pick-up trips. In contrast, if those requests would reserve, they are likely accepted as vehicles can reach them within the longer reservation period. Hence, such

[20] Actually, in the current implementation, the IRS is even more expensive than the optimization as the availability of vehicles has to be extrapolated far into the future for reservation requests.

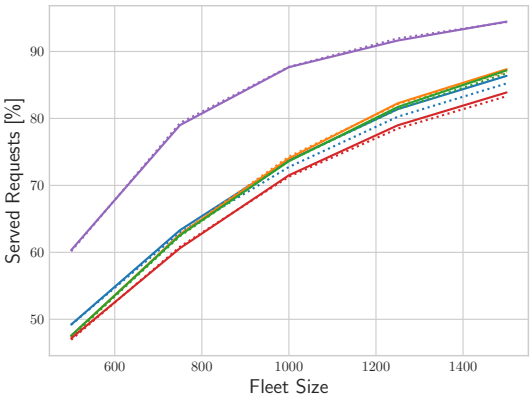


(a) Profit of hailing system.

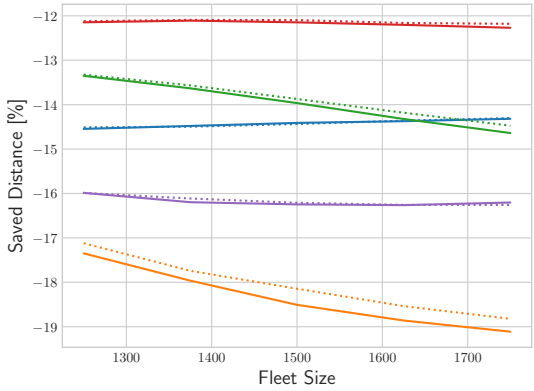
(b) Profit of pooling system.



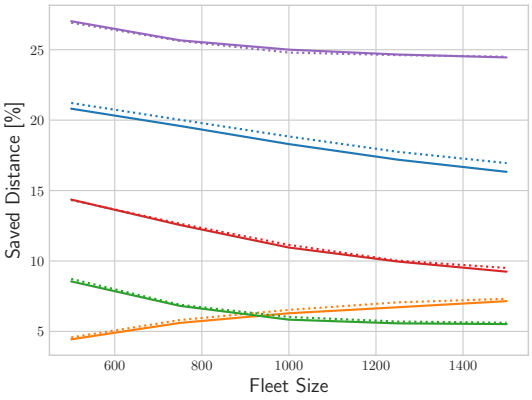
(c) Share of served requests for hailing system.



(d) Share of served requests for hailing system.



(e) Relative saved distance for pooling system.



(f) Relative saved distance for pooling system.

Figure 3.27: Evaluation of 5 % demand scenario with different shares of reservation requests.

System Type	Fleet Size	Reservation Share	Acceptance Rate of **Online** Requests	Acceptance Rate of **Reservation** Request
pooling	1500	25	83.3	100.0
		50	74.1	98.7
		75	62.3	90.4
hailing	1750	25	78.6	100.0
		50	65.5	99.7
		75	58.0	95.3

Table 3.5: Served requests per type (online/reservation) for selected scenarios with reservation.

requests will contribute to the empty miles but not the pooled distance, i.e., decrease the saved-distance indicator.

Table 3.5 gives more insight why the share of served requests does not increase compared to the online case. Most users with reservation can be accepted with a high probability as the fleet still has capacity in advance and a suitable vehicle can be found. In contrast, the share of online requests that can be accepted decreases with rising share of reservations since many vehicles are not available due to vehicle plans with reservation requests. Therefore, the effective density of available vehicles for online requests is much lower and the rejection rate much higher.

The total acceptance rate even decreases with the reservation share. This can likely be traced back to the selection of requests. While in the system with only online requests, the operator will serve those that are in the range of vehicles, some fleet vehicle is very probable to reach any reservation requests. This additional request constrains the possible routes that can be taken to include the online requests, thereby decreasing the likelihood of serving users more than in the less constraint online case.

In summary, an operation with reservation and online requests shows the same weakness as the service with only online requests: the vehicles need to be positioned according to demand forecasts in order to serve online requests with strict time constraints. This problem becomes even more apparent if part of the fleet is already utilized due to reservations.

3.7.3 Comparison of Repositioning Algorithms

The problem of balancing demand and supply in both temporal and spatial dimensions is addressed by repositioning. In this section, the results of different repositioning algorithms are compared for certain assignment strategies, namely the IRSBatch strategy with RV heuristics for hailing and the IRS strategy with RV heuristics for pooling. Additional to the repositioning algorithms introduced in Section 3.4.2, scenarios without repositioning and a widely used strategy based on [M. Pavone et al., 2012] are evaluated. Table 3.6 summarizes the different repositioning strategy scenarios, which are studied for varying fleet sizes respectively.

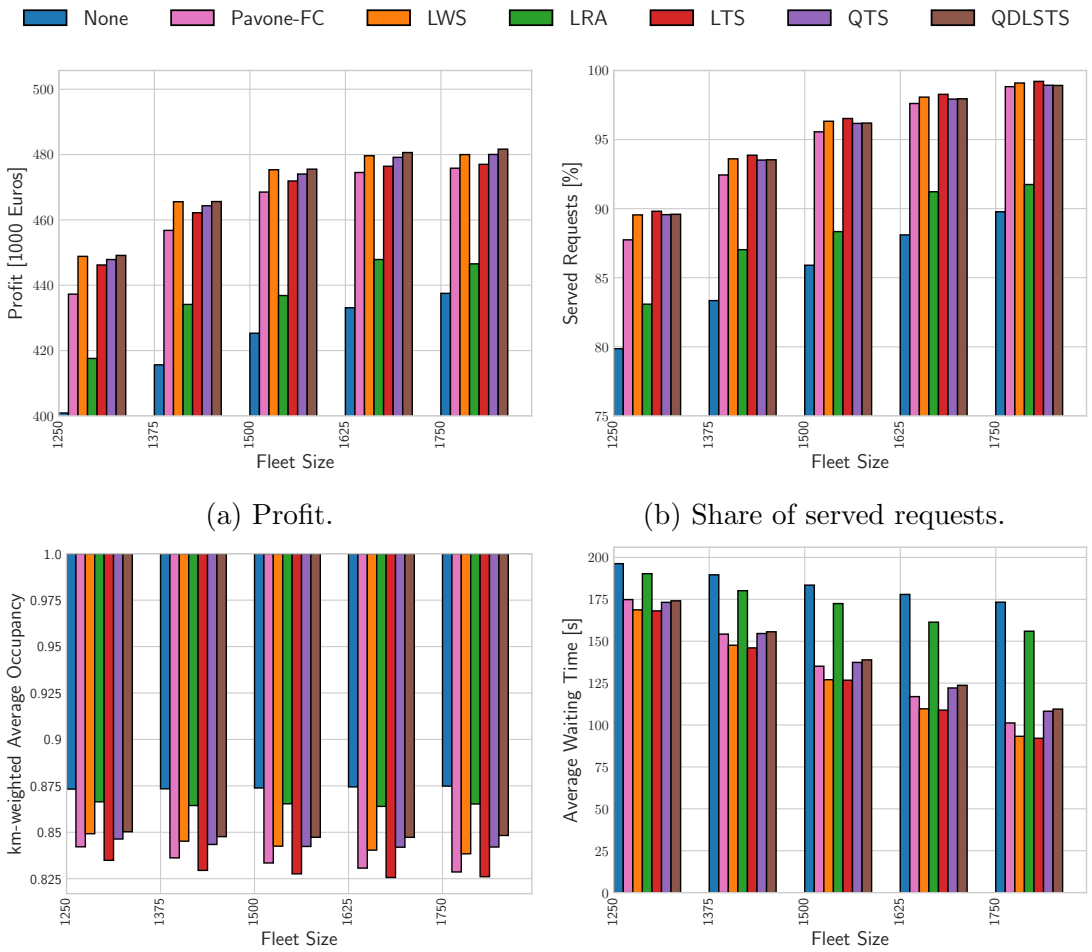
The impact of repositioning on the different assignment algorithms is studied in the subse-

Name	Zone / Density	Horizon Time-Steps	Objective	Other Comments
None	-	-	-	no repositioning strategy
Pavone-FC	zone	single	distance	see Appendix B.3 for details
LWS	density	multiple	linear weighted sum of deficit, stack, and distance	-
LRA	density	multiple	linear weighted sum of deficit, stack, and distance	reduced vehicle availability
LTS	density	multiple	priority: linear deficit and stack; secondary: distance	two-step solution approach
QTS	density	multiple	priority: quadratic deficit and stack; secondary: distance	two-step solution approach
QDLSTS	density	multiple	priority: quadratic deficit and linear stack; secondary: distance	two-step solution approach

Table 3.6: Repositioning strategy scenarios.

quent section. Moreover, the comparison of repositioning strategies assumes a certain set of hyper-parameters, namely a zone-to-zone correlation bandwidth of 1500 m and a future time horizon consisting of 4 steps with 15 min, and a certain zone system (see Fig. 3.15a). These inputs will be varied in sections 3.7.5 and 3.7.6.

Fig. 3.28 displays KPIs for the hailing case with varying fleet size and repositioning strategy. The benefit of repositioning becomes apparent from both operator and user perspective. The strategy derived from [M. Pavone et al., 2012] already improves the results considerably. Besides the LRA policy, all density-based multi-step strategies perform at least as good as the Pavone strategy. The LRA strategy underestimates the positive effects made by repositioning and therefore, less repositioning trips are performed. The share of empty repositioning mileage is only $1-2$ % for the LRA scenarios. The QDLSTS strategy generally performs best with operator profit increasing by $9-11$ % compared to the case without repositioning. The advantage decreases with rising fleet size, which is to be expected as the probability of a vehicle being in the right place increases with the fleet size. In contrast to the study without repositioning, the profit saturates around 1750 vehicles and the share of served requests reaches 99 %. As shown in Fig. 3.28b, the LTS formulation, which completely prioritizes the balance over expected repositioning distance, manages to serve most requests. However, this strategy also requires the most empty mileage and therefore has the lowest km-weighted average occupancy. The differences between the LWS and the quadratic formulations (QTS, QDLSTS) are rather small, especially for served requests. The stronger emphasis on the deficits in the QDLSTS increases the average occupancy slightly compared to the other two strategies making it the most efficient of the studied policies. The difference to the QTS strategy is a consequence of the QDLSTS strategy not sending vehicles between surplus zones. The reduction in empty repositioning distance compared to the LWS case is likely to originate from dynamic effects of



(a) Profit.

(b) Share of served requests.

(c) Average km-weighted occupancy.

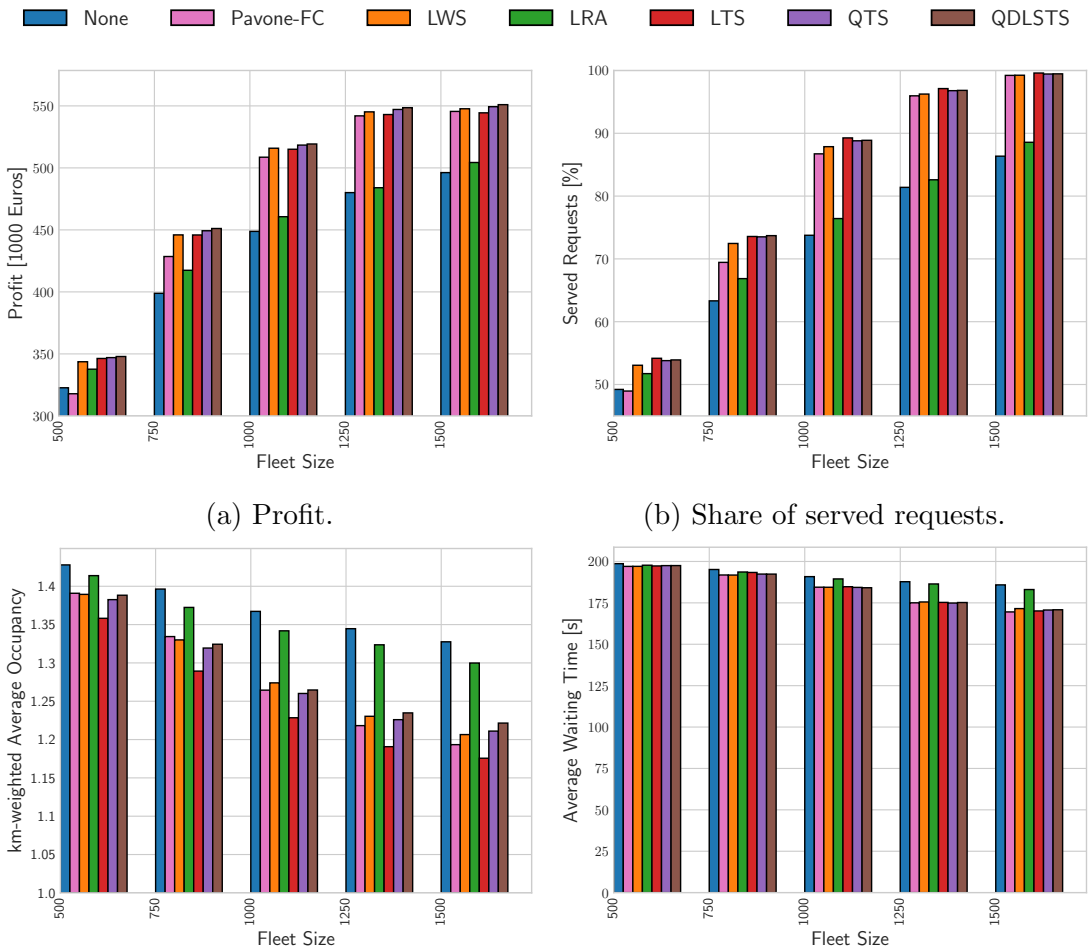
(d) Average waiting time.

Figure 3.28: KPIs for hailing system with 5 % demand scenario and different repositioning algorithms.

sending vehicles to the zones with larger deficits as the probability of such repositioning trips being not necessary is smaller. This is in line with the difference in empty mileage increasing with fleet size as the probability of unnecessary repositioning trips due to available vehicles increases. In general, the share of repositioning mileage can be kept below 10 percent. Even though more users are served, the repositioning strategies decrease the average waiting time (Fig. 3.28d).

The impact of the the tested forecast qualities is insignificant. The differences between scenarios with correct trip number per zone and time interval compared to scenarios with forecast errors are smaller for the QDLSTS strategy than for the LWS strategy, which can be attributed to the QDLSTS being less sensitive to small imbalances and thereby errors. Nevertheless, the resulting profits with and without forecast errors are within 1 % for all cases.

Very similar trends to the hailing results are observable for the pooling scenarios in Fig. 3.29. The profit saturates (here between 1250 and 1500 vehicles) as nearly 100 % of demand can be served with 1500 vehicles. Only the LRA strategy, which again is too passive, is an



(a) Profit.

(b) Share of served requests.

(c) Average km-weighted occupancy.

(d) Average waiting time.

Figure 3.29: KPIs for pooling system with 5 % demand scenario and different repositioning algorithms.

exception producing results similar to the no-repositioning case. The improvement in average user waiting time is much smaller than for the hailing case. For very small fleet sizes of 500 vehicles, the fleet is utilized by more than 70 % on average; therefore, there are hardly vehicles that can be repositioned and only 3 % of all VKT are for repositioning. Hence, it is clear that the impact of the different strategies is tiny for small fleet sizes. For the scenario with 1500 vehicles and the QDLSTS strategy, the daily average utilization is approximately 47 % and the share of repositioning VKT roughly 7 %. These additional empty trips are very effective though, as both the number of served customers is increased while the average waiting time is decreased. The cost of driving emptily is reflected in the km-weighted average occupancy, which is decreased from 1.33 to 1.22 for the 1500 vehicles scenario (from no repositioning to the QDLSTS strategy).

The comparison of the different strategies is also similar to the hailing case. The QDLSTS generates slightly more profit than the other strategies; the differences to the Pavone strategy are less pronounced than for the hailing case. As before, the Pavone strategy performs better

with higher vehicle number. The LTS strategy is again the most aggressive repositioning strategy, which serves most customers but requires most empty VKT and has the lowest occupancy.

The results suggest that exact spatial positioning is more important for hailing whereas routing combinations based on different vehicle distributions can still lead to similar overall results for pooling. Nevertheless, the results with repositioning show that the strategies also work for pooling scenarios. The forecast methodology, which assumes one vehicle is needed for one customer, obviously should have potential for improvement. This also becomes apparent from the results with forecast errors often producing better results than scenarios with perfect forecasts; the effect is not significant, though. The differences are below 1 % and therefore not illustrated. A simple pooling adaption, in which the incoming and outgoing expected trips were scaled down by zone-specific pooling rates, was tested but did not generate better results. Improving the utilization of forecasts, especially for pooling, remains an open research topic for future work.

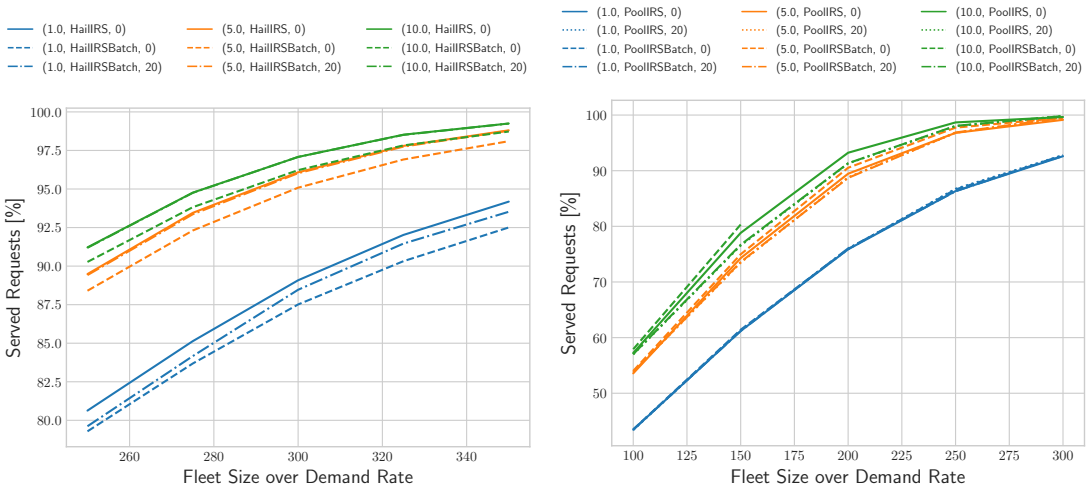
Since the number of simulations would become infeasible if all repositioning strategies were tested, operators will use the QDLSTS strategy, which performed best or among the best (with respect to profit) in the previous scenarios, for the remainder of this chapter.

3.7.4 Comparison of Assignment Algorithms with Repositioning

The choice of assignment algorithm is briefly revisited with the QDLSTS repositioning strategy. In general, the results of the comparison made in Section 3.7.1 repeat themselves, which is why only a part of the analysis is discussed here. The main difference is that the share of served requests and empty vehicle mileage are higher due to vehicle repositioning. However, explanations involving system dynamics become more complicated to follow and blurred as routes can be generated by either assignment, repositioning strategies or even both.

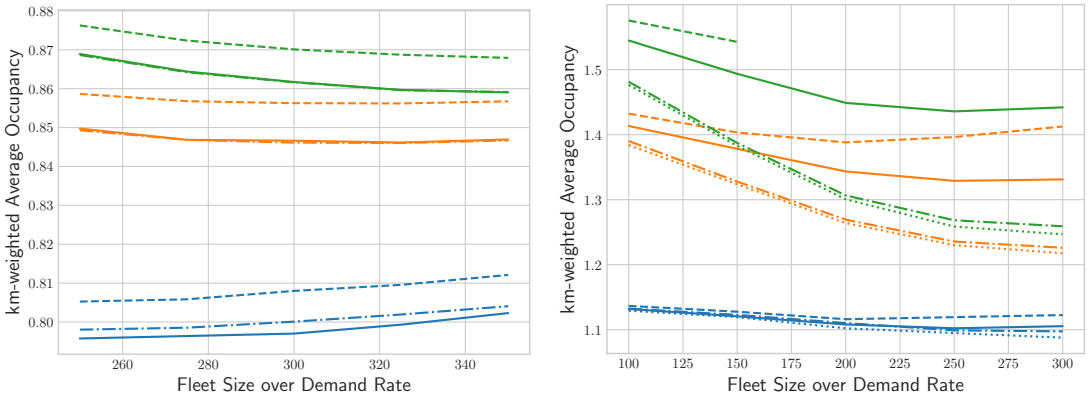
With repositioning, the scaling behavior becomes very apparent (see Figure 3.30). The more demand an AMoD system has (with scaled fleet size), the more customers can be served and the higher the average occupancy. The benefits of scaling demand from 1 % to 5 % are much more pronounced than those from scaling from 5 % to 10 %. Compared with this scaling behavior, the choice of assignment algorithm has a minor impacts on both of these KPIs. For all demand levels in the hailing case, the use of re-optimization still results in a trade-off with a lower share of requests being served but with a higher (km-weighted) average occupancy, i.e., less emptily driven kilometers. A higher optimization potential can be observed in the pooling case, where the re-optimization without RV heuristics is capable of serving more customers with a higher average occupancy. This comes at the cost of more computational effort, as the 10 % demand scenario cannot run in real-time on a single CPU in the given time threshold for larger fleet sizes. Similar to the case without repositioning, the RV heuristics decrease the average occupancy quite significantly, especially for larger ratios of fleet size over demand.

The daily profits are mostly determined by the demand level. Following the demand level, system type (hailing or pooling) and fleet size are the most relevant indicators. The repositioniong and assignment strategies only play minor roles. For hailing, the profits are in the ranges of $76 - 89$, $444 - 481$, and $919 - 976$ thousand € per day for the 1, 5, and 10 % demand levels, respectively. Similarly, these ranges are $50 - 100$, $346 - 575$, and $951 - 1162$



(a) Share of served requests for hailing scenarios.

(b) Share of served requests for pooling scenarios.



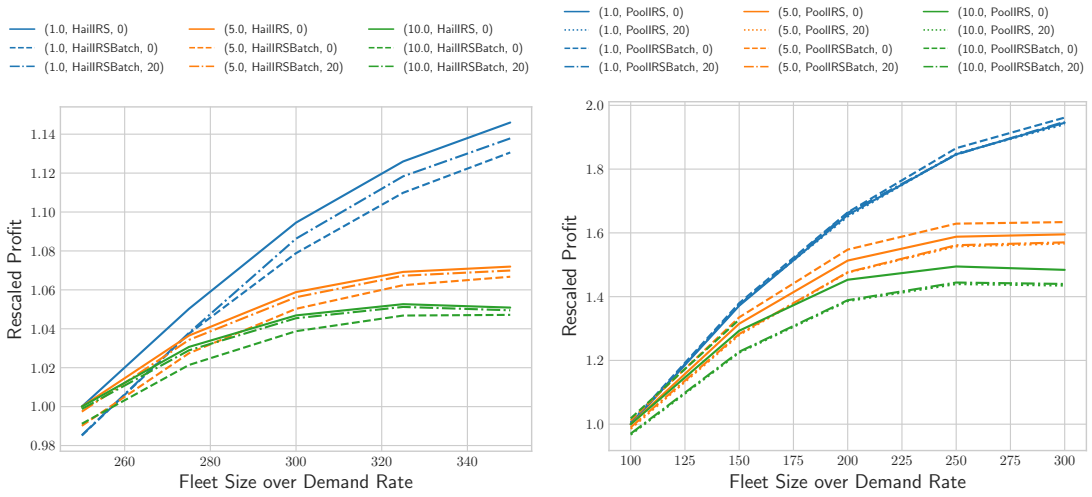
(c) km-weighted average occupancy for hailing scenarios.

(d) km-weighted average occupancy for pooling scenarios.

Figure 3.30: Share of served requests and average occupancy for variations of demand, fleet size, and assignment strategy for scenarios with QDLSTS repositioning. The tuples in the legend refer to (i) demand level, (ii) assignment with or without re-optimization, and (iii) applied RV heuristics, where 20 is the number of vehicles selected by the RV heuristics and 0 means no RV heuristics are used.

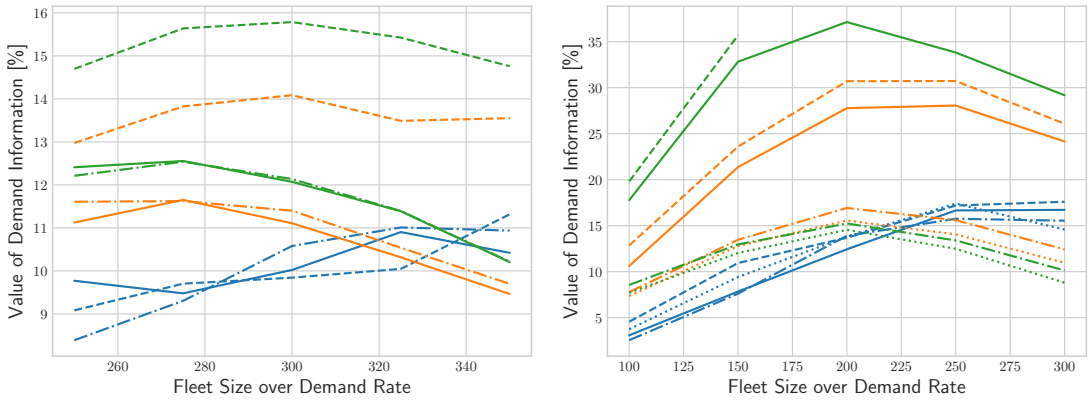
thousand € per day for the pooling scenarios.

To illustrate the dependence on fleet size and assignment strategies, Figures 3.31a and 3.31b show a rescaled profit, where the profit of a scenario is divided by the profit of the scenario with HailIRS/PoolIRS without RV heuristics and the lowest tested fleet size for the respective demand levels. It can be observed that the dependence on fleet size is strongest for small demand levels. It should be noted that the plot already scales fleet size with demand level, so the effect is more than linear. These statements hold true for both hailing and pooling. The larger values for pooling are a consequence of the larger fleet size range analyzed in this case



(a) Rescaled profit: comparison with HailIRS scenarios for 250 vehicles per demand level.

(b) Rescaled profit: comparison with PoolIRS scenarios for 100 vehicles per demand level.

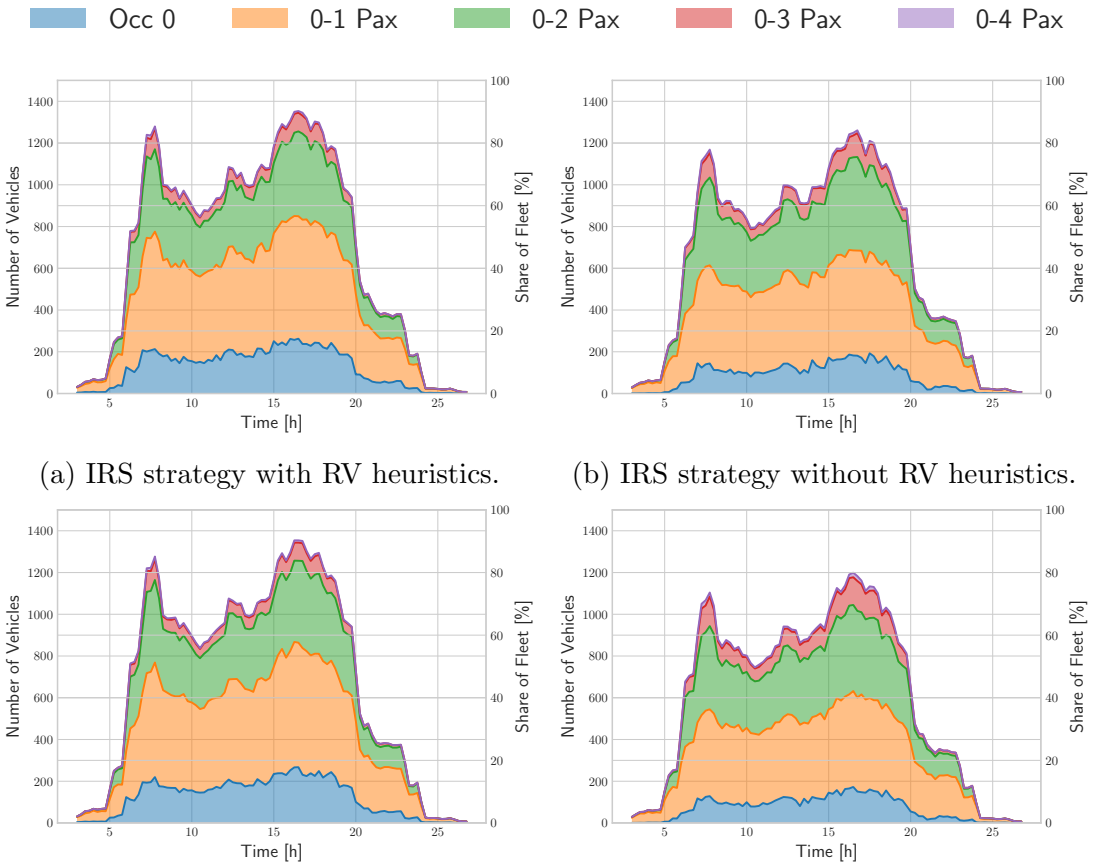


(c) Value of Information for hailing scenarios.

(d) Value of Information for pooling scenarios.

Figure 3.31: Evaluation of profit for variations of demand, fleet size, and assignment strategy for scenarios with QDLSTS repositioning. The tuples in the legend refer to (i) demand level, (ii) assignment with or without re-optimization, and (iii) applied RV heuristics, where 20 is the number of vehicles selected by the RV heuristics and 0 means no RV heuristics are used.

study. As the marginal increase in profit with fleet sizes is approximately 0 for the 5 % and 10 % demand scenarios, the optimal fleet size is likely part of the studied ranges or slightly above. Only the 1 % scenario would really benefit from more vehicles in the fleet. Even with repositioning, the re-optimization of request-vehicle assignments does not generate additional profits with the selected control function as shown in Figure 3.31a. The optimization with RV heuristics pruning the solution space even produces better results, which is a clear indicator that the control objective, which should encompass the dynamic effects in a static problem, could be chosen better. As the re-optimization tends to shift assignments to save VKT, it



(a) IRS strategy with RV heuristics.

(b) IRS strategy without RV heuristics.

(c) IRSBatch strategy with RV heuristics.

(d) IRSBatch strategy without RV heuristics.

Figure 3.32: Stack plots showing vehicles with certain occupancy for pooling scenarios with 5 % demand and 1500 vehicles and the QDLSTS repositioning strategy. The 0-occupancy vehicles are divided into empty trips (blue) and idle vehicles, which are represented by the white area on top.

would likely be beneficial to put more weight on the system time of an assignment (c^{VOT}) over the distance-related component (c^{D}) in the hailing control objective in equation (3.9). For pooling operations, the re-optimization brings clear benefits, at least from a certain scale of the system but only without the RV heuristics that severely limit the solution space. For small scale systems and small fleet sizes, the heuristics are practically not limiting the solution space and have hardly any impact.

Next, a comparison between the results with and without repositioning is conducted for the different assignment strategies. To this end, the *value of demand information* [Wen, Nassir, et al., 2019] is defined by

$$\text{Value of Demand Information} = \frac{P' - P}{P} \tag{3.69}$$

where P' and P denote the profit with and without repositioning, respectively. This value

is defined for each scenario (demand level, fleet size, assignment strategy) independently. Figures 3.31c and 3.31d show some similar trends for hailing and pooling. In both cases the scenarios with full re-optimization, i.e., re-optimization and no RV heuristics, have the highest value of demand information. This is likely a consequence of the dynamic effects sending (comparatively) more vehicles to zones where they are not needed in scenarios without repositioning, which are moved by the repositioning strategies. Additionally, the dependency on fleet size results in a convex curve in the studied range in most cases except for the 1 % demand scenario. This is reasonable because the probability of finding an idle vehicle to reposition increases with fleet size and the probability of a vehicle deficiency to occur decreases with fleet size. Therefore, the the highest benefit of repositioning can be expected for a medium fleet size, for which approximately $75-85$ % and $90-95$ % of requests can be served without and with repositioning, respectively. As the 1 % scenario needs more vehicles to reach these numbers, it can be speculated that the peak of value of demand information is reached for fleet sizes above the studied range. Interestingly, the value of demand information is significantly higher for the pooling than the hailing scenarios (both with re-optimization). This is not necessarily intuitive as the proximity of empty vehicles should be less important for a pooling operation, where vehicles with customers on-board can still be considered for assignment. However, in a pooling system an additional vehicle at the right place can serve more than just one additional request. At least for the scenarios with re-optimization, in which the average occupancy is higher, this effect seems to prevail in the studied scenarios.

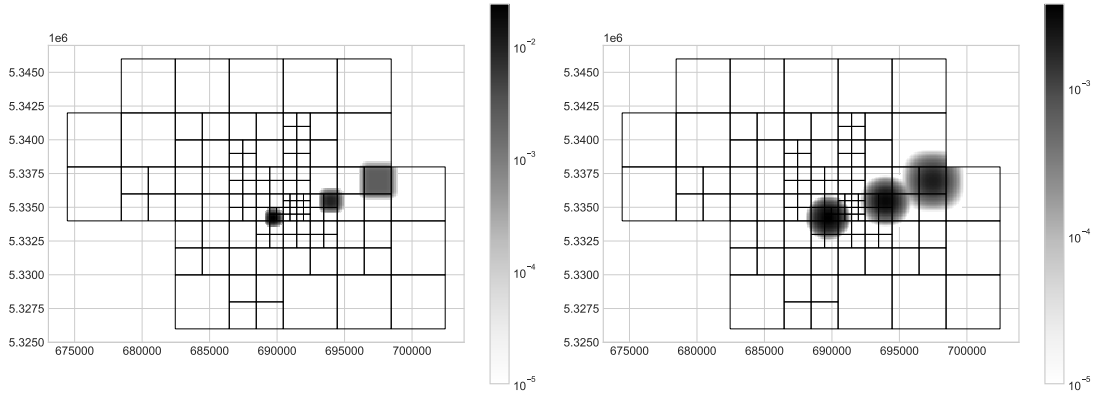
Evaluating the utilization and occupancy for a pooling system over time shows clear differences without (see Figure 3.22) and with repositioning (see Figure 3.32). Unlike in the case without repositioning, the utilization curve now follows the demand curve as very high shares of demand can be served throughout the day. For the 5 % demand scale illustrated in Figure 3.32, it is still observable that most trips have 1 or 2 requests on board. The comparison of occupancy levels for the different assignment strategies shows that with RV heuristics, more vehicles are in operation – approximately 90 % with versus 80 % without in the peak – where most of the additional vehicles have no or only one passenger on board. Smaller efficiency gains in the range of 5 % of vehicles become observable when applying batch re-optimization without RV heuristics. With RV heuristics, re-optimization has negligible impacts on the occupancy.

As the computation time for a re-optimization is quite significant for pooling simulations, the remainder of this chapter uses the HailIRSBatch and PoolIRS strategies with RV heuristics.

3.7.5 Sensitivity Analysis of Repositioning Hyper-Parameters

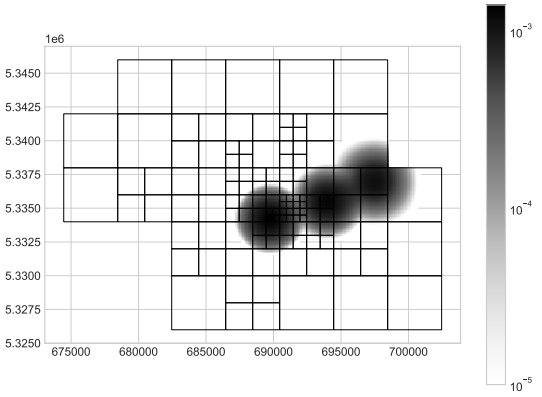
The density-based repositioning strategies have two hyper-parameters that tune the algorithms: the number of forecast steps J and the bandwidth h, which describes how far the availability of vehicles spreads from a single point and thereby determines correlations between zones. To clarify this effect, Figure 3.33 illustrates the imbalance density for various bandwidth parameters, where the imbalance value is set to 1 for three zones of different size and 0 for all others. It can be seen that the correlations of neighboring zones are more relevant (1) for larger values of h and (2) the smaller the zone size.

Additionally, the QDLSTS has a parameter N^{esq} determining the weight between the



(a) Bandwidth $h = 500$ m.

(b) Bandwidth $h = 1500$ m.



(c) Bandwidth $h = 2500$ m.

Figure 3.33: Example of imbalance density plots for various bandwidth parameters. The imbalance values of three zones are set to 1.

quadratic deficit and the linear stack terms in equation (3.52). For this case study, $N^{esq} = 10$ was set based on the intuition that the quadratic term should become more important if a deficit is larger than 10 vehicles. Simulations with $N^{esq} = 5$ and $N^{esq} = 20$ were also conducted but did not bring much insight. The KPIs were hardly affected with relative changes below 0.1 % in the hailing and below 0.5 % in the pooling case and without clear trends. Hence, these results are not illustrated and the analysis focuses on the hyper-parameters existing for all repositioning strategies.

In general, hyper-parameters can have a huge impact on the performance of algorithms. However, only minor impacts can be observed in this case study, also for bandwidth h and number of forecast steps J. In all scenarios, a variation in fleet size has much higher effects. To visualize the impacts of the other two hyper-parameters, the fleet size was kept constant in Figure 3.34 and the y-axes of the figures are focused on the differences. The fleet size was chosen to serve approximately 96 % of demand for both hailing and pooling. The tested variation of bandwidth has a higher impact than the variation of the number of forecast steps. 2, 4, and 6 were chosen as the number of forecast steps, which translate into total forecast time horizons of 30 minutes, 1 hour, and 1.5 hours, respectively. The choice of 6 steps tends

(a) Profit in hailing system.

(b) Profit in pooling system.

(c) Served requests in hailing system.

(d) Served requests in pooling system.

(e) Saved distance in hailing system.

(f) Saved distance in pooling system.

Figure 3.34: KPIs for variation of repositioning hyper-parameters with a fleet size of 1500 (hailing) and 1250 (pooling) vehicles.

to perform slightly worse with respect to profit as it generates additional empty mileage.

Compared to the hardly observable changes by the variation of the forecast horizon, it can be noticed that the introduction of the bandwidth, i.e., the step from bandwidth 1 meter to 500 meters, results in the largest deviations in the results. Even these deviations are rather small: the profit changes by approximately 3500 € per day, which is less than 1 % of the daily profit. Nevertheless this introduction of the bandwidth is interesting as with the introduction of the bandwidth both components of the objective, i.e., the revenue from served customers and the costs from driving can be improved, for hailing and pooling. Additionally, this difference in daily profit accumulates to a sizable sum of more than a million € over a year.

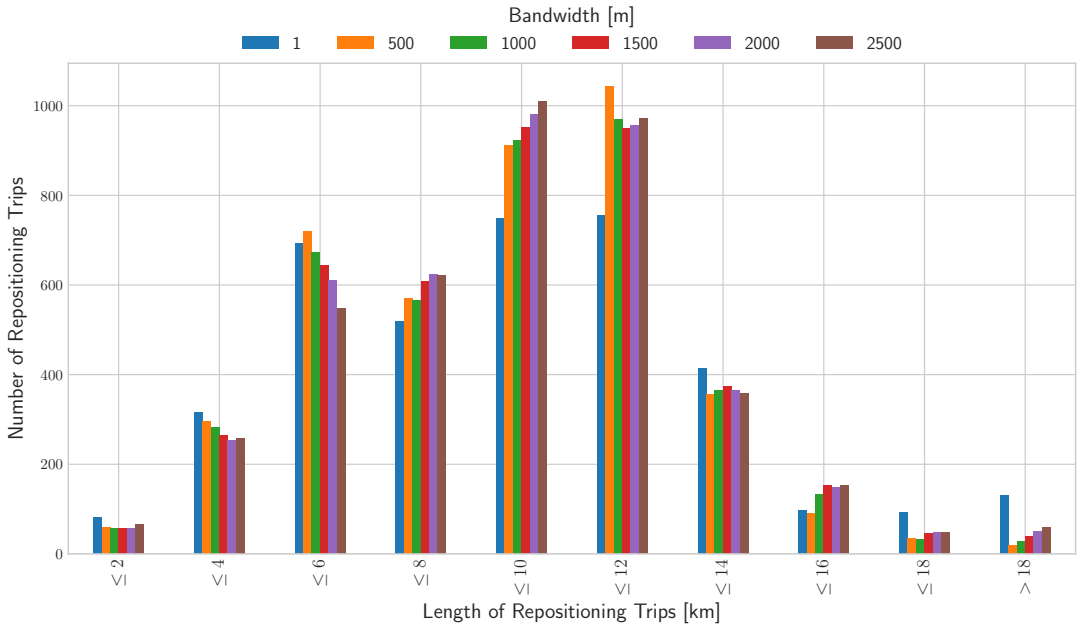


Figure 3.35: Trip length distribution for hailing scenarios with fleet a fleet size of 1250 vehicles, four forecast steps, and varying bandwidth parameter h.

All other steps in bandwidth result in almost negligible differences. Bandwidth parameters between 500 meters and 1500 meters create a trade-off for the hailing case. On the one hand, increasing the bandwidth slightly increases the share of served requests (Figure 3.34c) but also slightly increases the share of empty fleet mileage, which is equivalent with decreasing the saved distance KPI (Figure 3.34e). Higher bandwidth parameters of 2000 meters and 2500 meters perform slightly worse in both categories. For pooling, the bandwidth of 500 meters performs slightly worse and the larger bandwidth parameters slightly better than in the hailing case. Figure 3.34f still shows the same trend with saved distance decreasing with increasing bandwidth.

The saved distance results can be traced back to changes in the share of distance driven for repositioning. To get further insight, the repositioning trip length distribution is illustrated in Figure 3.35 (for a hailing scenario). With bin size of 2 km, the largest share of repositioning trips is in the range between $8 - 12$ km. Longer trips occur less, which is reasonable as the methodology aims at repositioning trips that will finish within the first forecast step of 15

minutes. In this time, vehicles with an average speed of 48 km/h can drive 12 km. As before, the choice of no bandwidth ($h = 1$ m) creates a clearly distinct picture: the peak between $8 - 12$ km is much less pronounced and instead more trips with higher distance are scheduled. These trips are generally using the faster motorway and arterials to move vehicles from the outside of the operating area into the center. As the deficits are not spread among neighboring zones and the algorithm tries to equalize deficits of the individual zones, the destination of repositioning trips is often further in the center resulting in longer trips. The differences in trip length distribution for $h \geq 500$ m are difficult to interpret. As the vehicle availability is also spread with the bandwidth, it is likely necessary to make more shorter trips (below 6 km) to fine-balance the zones when the bandwidth is smaller.

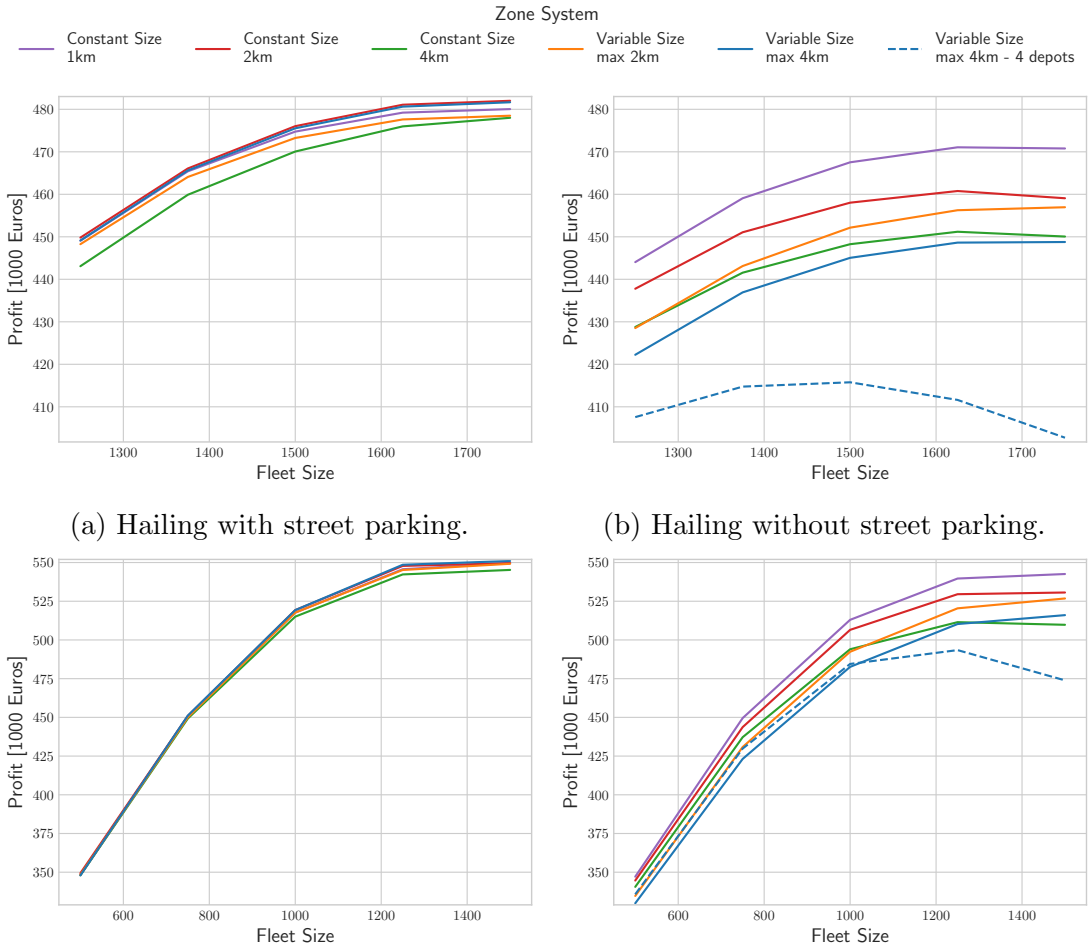
It can be summarized that the hyper-parameters of the density-based QDLSTS approach have minor impacts in the case study and the selected parameters of $h = 1500$ m and $J = 4$ are appropriate choices for the remainder of this chapter.

3.7.6 Analysis of Different Zone-Systems and Parking Possibilities

In this subsection, different zone systems are utilized for the demand forecasts and repositioning. Moreover, scenarios are investigated, where the operator is not allowed to park on public parking lots and vehicles have to return to depots of the operator, which are assumed in the center node of each zone. This parking prohibition is studied as German law requires taxi and private-sector hailing and pooling companies to keep to the *Rückkehrpflicht*, which means that vehicles are not allowed to remain idle and wait for customers in a zone if their original base is in another zone. This regulation inherently creates approximately 50 % empty travel. Administrations can allow additional places to park, which is modeled here with the operator depots. Additionally to the scenarios with depots in the zone centers, scenarios are studied, in which there are only four depots positioned near the Main, East, and Pasing train stations and Parkhaus Schwabing, respectively. If fleet vehicles can only park in depots, the operator repositions the AVs towards the nearest depot whenever they become idle after dropping of the last on-board customer without having new pick-up requests assigned to their route. At any time during their trip to the depot, these vehicles can be assigned to new requests and they directly drive towards the new request, i.e., they do not have to finish driving to the depot first. Besides this parking limitation, the scenarios are defined by the default parameters given in Table 3.3. In the following, scenarios with parking limitations are denoted by *without street parking*. The others are denoted by street parking, where it is not further specified if this is on-street or on parking bays.

Figure 3.36 displays the profits the operators generate in the studied scenarios. Comparing scenarios with street parking and different forecast zones (Figures 3.36a and 3.36c), results show that all studied zone systems perform similarly. Only the zone system with the least spatial resolution, which is a zone system with a constant size and edge length of 4 km, performs considerably worse. Other than expected, the profit does not grow monotonously with spatial resolution. Fig. 3.37 illustrates that a higher resolution of zones performs better with respect to empty VKT, however, at the cost of serving less users.

In contrast, the picture looks very different for the scenarios without street parking. The first observation is that profits in the hailing operation are significantly smaller. Depending



(a) Hailing with street parking.

(b) Hailing without street parking.

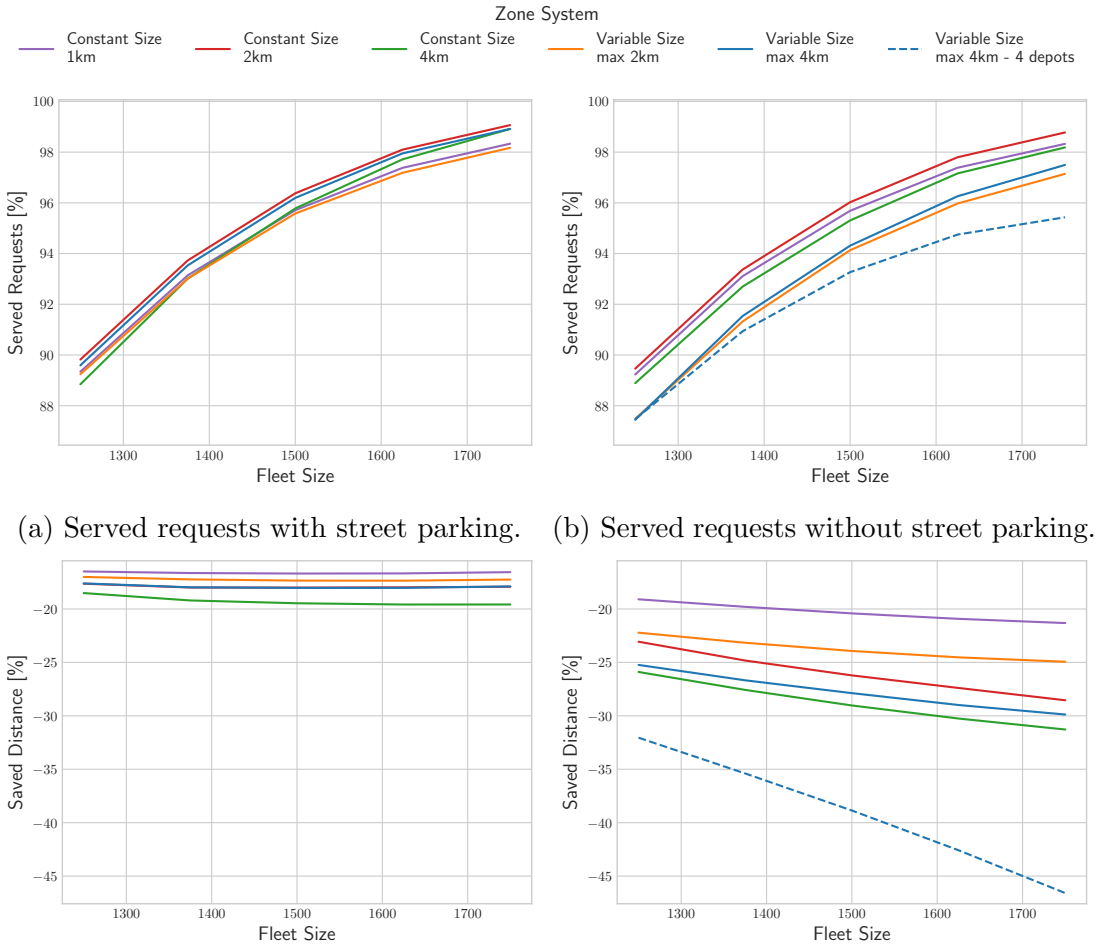
(c) Pooling with street parking.

(d) Pooling without street parking.

Figure 3.36: Profit for scenarios with different zone systems and street-parking possibility.

on zone system and fleet size, the return-to-depot requirement generates between $10 - 70$ thousand € per day smaller profits than the respective street parking scenario. The zone systems with higher spatial resolution perform better because of the higher proximity of the nearest depot. For the same reason, the scenarios with only four depots is by far the worst from an operator's point of view and the effect becomes even worse with increasing fleet size. In the pooling scenarios illustrated in Figure 3.36d, the same trends are observable. As the range of studied fleet sizes in the pooling scenarios start from a lower value, the differences are not as noticeable at first because the fleets are often in an under-supply state, in which the vehicles do not become idle.

From a customer's perspective, the rates of served/rejected requests is the most important indicator. As shown in figures 3.37a and 3.37b, the effect of the street parking restrictions on the share of served requests is rather low. Since in times of high demand vehicles are unlikely to become idle, the number of return trips to depots are limited. Nevertheless, if there are only four depots, slight losses are observable because repositioning vehicles have to return to depot if they are not matched to a request near their destination zone. If a vehicle does



(a) Served requests with street parking.

(b) Served requests without street parking.

(c) Saved distance with street parking.

(d) Saved distance without street parking.

Figure 3.37: Share of served requests and saved distance for hailing scenarios with different zone systems and street-parking possibility.

not find a match on its return trip, it is likely redistributed again. This behavior obviously generates a significant amount of empty travel, especially for larger fleet sizes. Therefore, this scenario creates up to 45 % of additional driven distance. Compared to this large value, the additional $5 - 15$ % empty VKT in the other scenarios without street parking seem quite small even though this increase is quite significant as well. As mentioned before, the proximity of the nearest depot is decisive for the respective increase in empty VKT and consequently the decrease in saved distance. With street parking, the share of empty travel also follows the same order: the smaller the zones, the less empty travel is conducted. The zones with variable sizes approximately generate as much empty VKT as the zone system with half of their maximum edge length. With respect to served requests, differences are within a few percent for all scenarios, even for the one with only four depots. The ability to add new requests to existing routes can make up for the inefficient cruising of vehicles. However, the statements for empty VKT are also true for a pooling system.

The costs of having a zone system with very high spatial resolution have to be considered

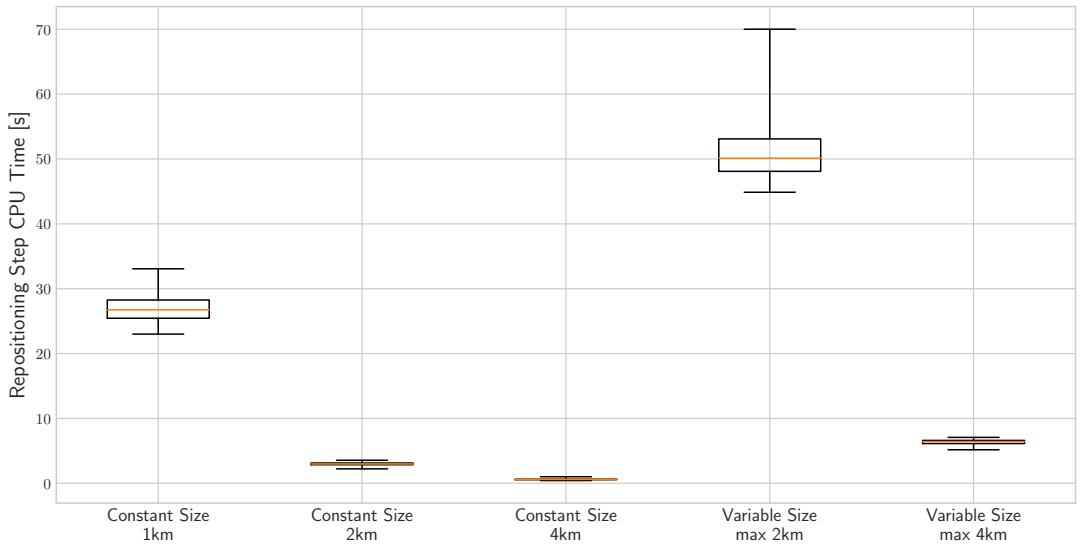


Figure 3.38: Computation time of repositioning steps per zone system. The values reflect the mean values of all scenarios with the respective zone systems and 1625 hailing and 1500 pooling vehicles.

when choosing a zone system. The main cost is the computational effort that is generated by a larger number of zones. As the studied repositioning algorithms are zone-based, fleet size hardly affects the computation time of single problem instances. Moreover, it does not matter whether the system is operated in hailing or pooling mode. Therefore, Figure 3.38 summarizes the computation times of all studied scenarios with the respective zone system. It is apparent that the computation time for single repositioning steps becomes higher for zone systems with higher resolution. The non-linear increase can be expected as the number of zones grows by a factor of 4 when decreasing the edge length to half its size and the number of decision variables is quadratic in the number of zones. For a real-world repositioning algorithm, computation times of approximately 30 or 50 seconds bring additional difficulties as a synchronization issues between the computed repositioning solution and the new system state at the time the solution is available can be expected to grow with the computation time.

3.7.7 Analysis of Dynamic Pricing Strategies and Forecast Quality

For this part of the study, it is assumed that the sensitivity of requests with respect to dynamic pricing (DP) follows the logistic function in equation (3.59) with $a = b = 1$. Additionally, it is assumed that all offers with dynamic pricing factor 1 are accepted. To this end, the requests are modeled with a maximum acceptable pricing factor, which is drawn from the distribution

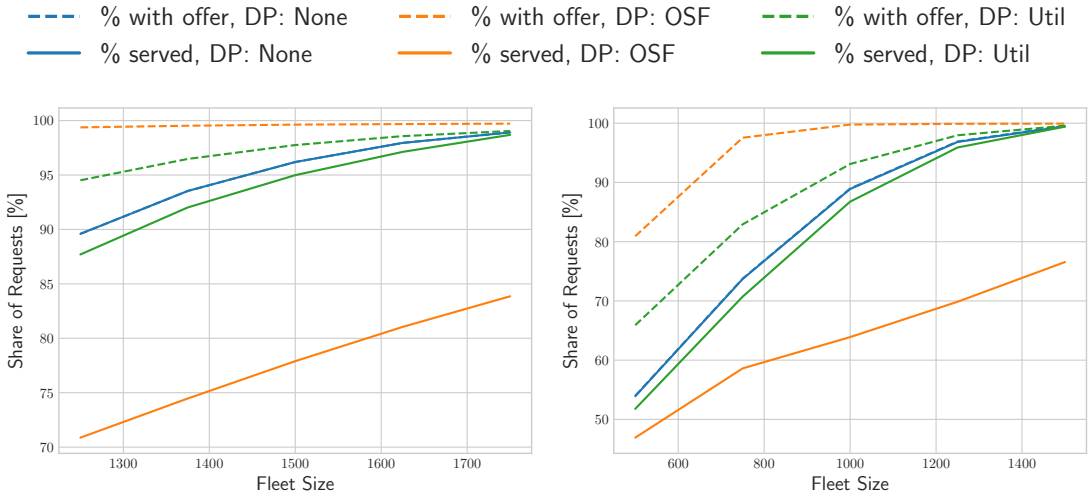
$$\Omega(\beta^p) = -\frac{dD(\beta^p)}{d\beta^p} \tag{3.70}$$

where $\beta^p \in [1, \infty[$ and the logistic function $D(\beta^p)$ in equation (3.59) is normalized to satisfy $D(1) = 1$. Then, the probability of accepting an offer with dynamic pricing factor $\mathcal{P}(\beta^p)$ is

given by

$$\mathcal{P}(\beta^p) = 1 - \int_1^{\beta^p} \Omega(x)dx = 1 - (D(1) - D(\beta^p)) = D(\beta^p) \tag{3.71}$$

The operator also uses $a = b = 1$ in its determination of the dynamic prize in the one-step forecast (OSF) based dynamic pricing strategy. The maximum surge factor $\tilde{\beta}^p = 2$ for the OSF strategy. To receive the same maximum factor in the utilization based strategy (Util), $\beta^{PU} = 4$. In this thesis, the original goals of the strategies is that (i) most requests receive a feasible offer but (ii) a similar amount of users reject as the operator rejects in the case without dynamic pricing. Hence, the price point is not the result of a profit optimization where less users would be served albeit for a higher fare.



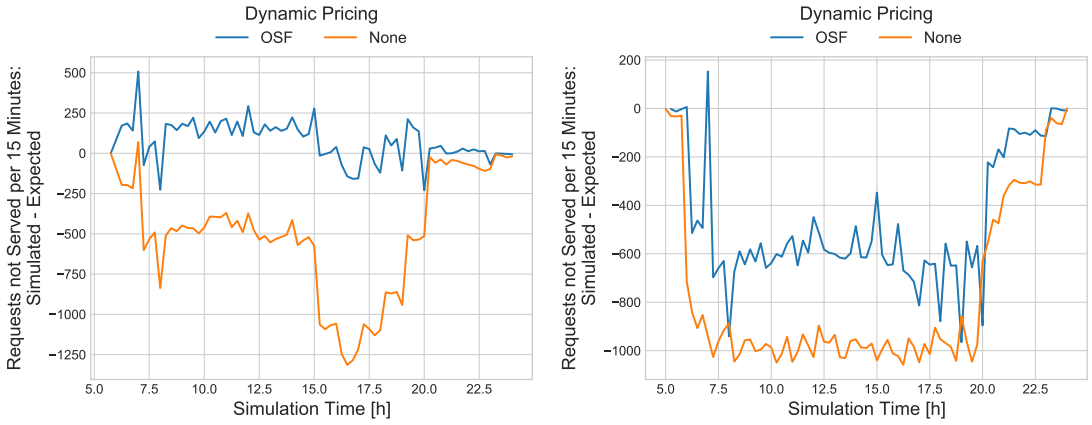
(a) Hailing service with 1250 vehicles.

(b) Pooling service with 750 vehicles.

Figure 3.39: Share of requests that (i) received an offer and (ii) were served without and with dynamic pricing.

It can be observed in Figure 3.39 that this goal is not achieved properly by the OSF pricing strategy. The share of served requests is significantly lower than in the scenario without dynamic pricing. The figure also shows curves for the share of users that received an offer of the operator. It can be derived that the fares are set so high that "too many" users reject their offers in the OSF strategy. At least the goal of making feasible offers to every request is fulfilled for hailing. With less requests accepting offers, the fleet is less utilized and therefore new requests can be inserted into the vehicle plans within their time constraints. In the pooling case with 500 vehicles, the fleet is simply to small to even achieve that with a maximum pricing factor of $\tilde{\beta}^p = 2$, which is set in most zones for most of the time between 6am and 8pm. In general, the utilization based dynamic pricing strategy works quite well: the share of served requests is very similar and most requests receive an offer. It could set a little higher fares in the pooling scenarios with small fleet sizes to increase the share of requests that receive offers.

Figure 3.40 illustrates the reason why the OSF strategy does not work as intended. For the hailing case (Figure 3.40a), the OSF strategy actually does what it is supposed to quite well: throughout the day, it uses forecasts to set the price point in a way that all requests can



(a) Hailing service with 1250 vehicles. (b) Pooling service with 750 vehicles.

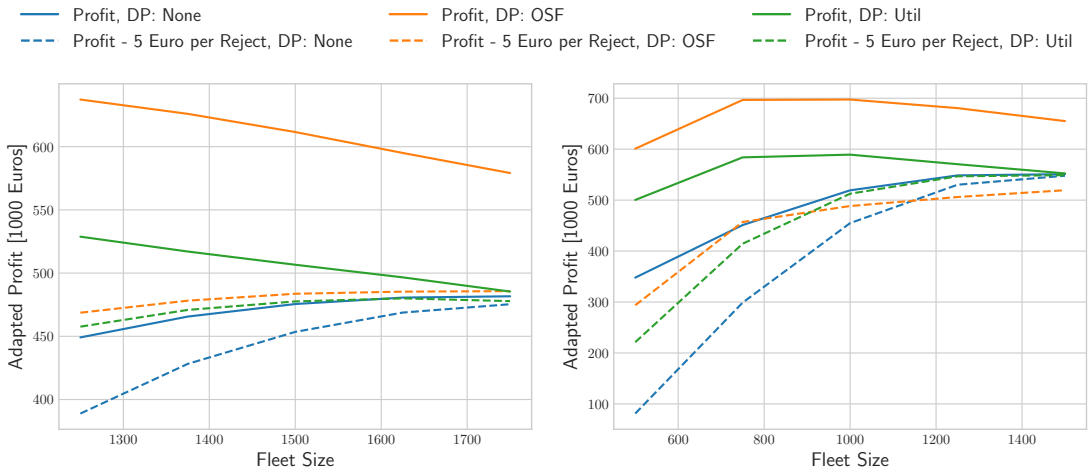
Figure 3.40: Comparison of forecast of expected vehicle deficits of next 15 minutes and requests not served in the simulation of these 15 minutes without and with dynamic pricing.

receive an offer. Indeed the forecasts of requests that will not be served — either due to users not accepting a too high price or the operator not having sufficient supply — is accurate and matches the share of requests that are not served in the next 15 minutes of the simulation quite well, i.e., the blue curve in Fig. 3.40a is close to 0. However, this is mostly due to the fact that the share of requests not accepting a certain fare is accurate.

Simulations without dynamic pricing show that the imbalance forecast I_z^1 from equation (3.24) actually has quite some room for improvement as the available supply is constantly underestimated. The reason for this underestimation is that vehicles are only counted for the region they are in or will end up due to their current assignments; therefore, the supply side estimation is bound from the top by the fleet size. However, the number of trips that can be conducted and are predicted by the demand forecast can be higher than the fleet size as multiple trips with shorter duration can be made within the forecast time horizon of 15 minutes.

All these statements are true for the pooling case as well, but there is an extra effect creating the larger gap visible in Figure 3.40b. The forecast of supply should not only consider the number of vehicles but also the number of seats. Furthermore, to increase difficulty for the pooling imbalance forecast, the OD relations and therefore the direction in which seats are available become relevant. As the used forecasts show very good results in the repositioning context, an improvement of the imbalance forecasts are considered out of the scope of this thesis and remain a future research topic.

As a consequence of the underestimation of supply, the OSF strategy sets the dynamic pricing factors higher than the utilization based strategy. Higher dynamic pricing factors would also be the result of optimizing the expected profit as the expected profit can be interpreted as a convolution of the dynamic pricing factor and the probability of users accepting a certain pricing factor. Hence, the full lines in Figure 3.41 show the highest profits being generated by the OSF strategy. Since users likely value a high service rate and might make less service requests if they are often rejected or too expensive, opportunity costs are considered in Figure 3.41



(a) Hailing service.

(b) Pooling service.

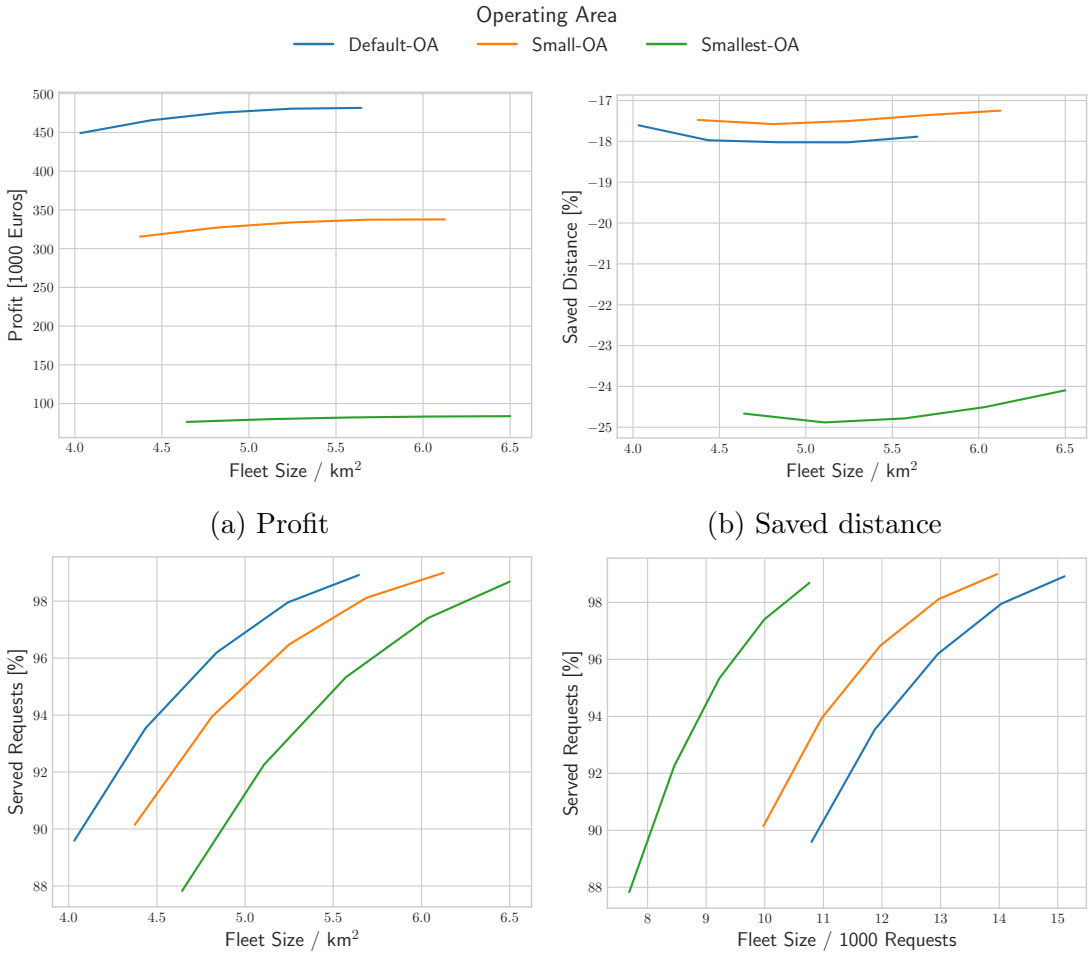
Figure 3.41: Profits and adapted profits including opportunity costs without and with dynamic pricing.

with the dashed-line curves. For this thesis, a value of 5 € per rejection is assumed; typically this value would be derived from market research. With the opportunity costs part of the dynamic pricing objective, the adapted profit optimization is likely to have a similar effect as the dynamic pricing strategies in this thesis, namely to set fares in a way that as many requests reject as would have to be declined due to missing supply.

3.7.8 Choice of Operating Area

In this section, the three operating areas illustrated in Figures 3.15a, 3.16a, and 3.16b. For this experiment, the fleet sizes are approximately scaled with the area of the respective operating areas. It can be noted that the smallest operating area has a very high similarity to the business area of today's carsharing providers in Munich. For the largest operating area, which is denoted by "Default-OA", the fleet sizes are between 1250 and 1750 for hailing and between 500 and 1500 for pooling, just as before. For the next operating area denoted by "Small-OA", the respective ranges are from 910 to 1274 for hailing and from 364 to 1092 for pooling. For the last operating area, the "Smallest-OA", only $260 - 364$ and $104 - 312$ vehicles are used for hailing and pooling, respectively. The ratios of fleet size over area are a little larger for the smaller operating areas. This was a result of choosing an integer interval between the five fleet size experiments.

The evaluation starts with a hailing service again. The computation time for the hailing scenarios with re-optimization follows the area sizes. The Default-OA scenarios last approximately 4.6 h, the Small-OA scenarios approximately 2.8 h, and the Smallest-OA scenarios approximately 0.7 h. The evaluation of profits is shown in Figure 3.42a. It is a good sign that profits are increasing with the size of the operating area as it is profitable for AMoD operators to extend their service area beyond the city center, where PT infrastructure is anyway well developed. Of course, this result is valid for the assumed demand, cost, and fare structure.



(a) Profit

(b) Saved distance

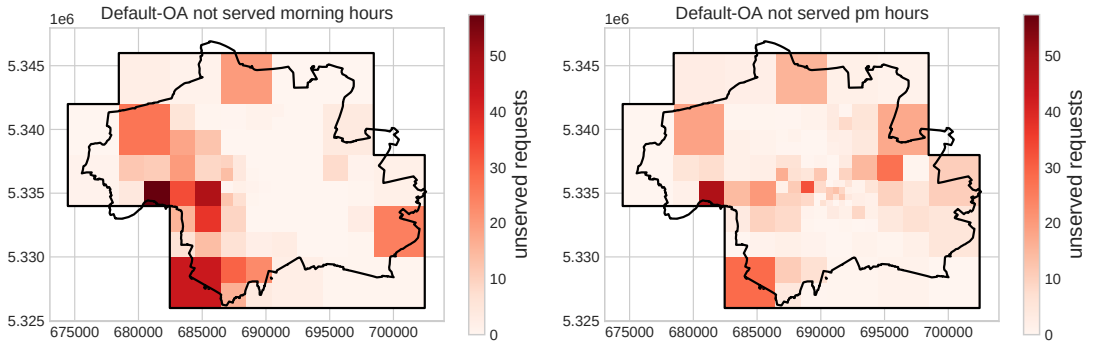
(c) Share of served requests with fleet size scaled per area.

(d) Share of served requests with fleet size scaled per 1000 requests.

Figure 3.42: KPIs for hailing service for various fleet sizes in different operating areas.

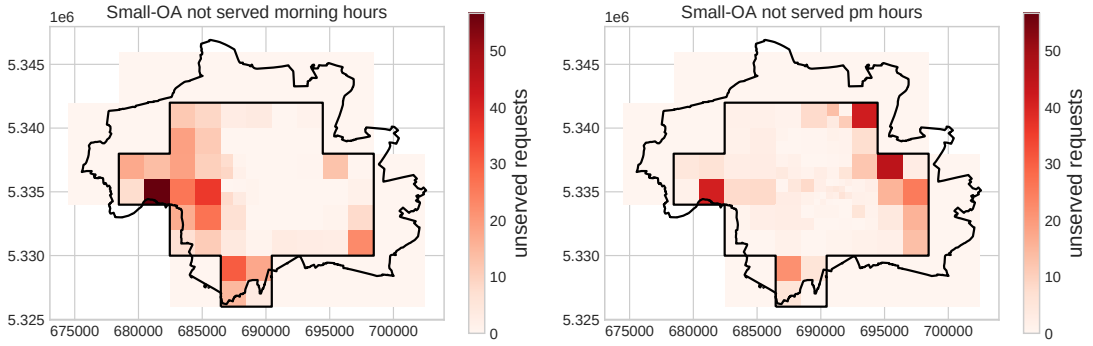
As illustrated in Figure 3.42b, the hailing systems generates around 18 % of additional VKT in the two larger scenarios, whereas approximately 25 % of additional VKT is created in the Smallest-OA scenarios. This result might seem a little surprising as more repositioning VKT is required for the larger operating areas ($3-5$ %) than in the Smallest-OA scenario ($1.5-3$ %). Nevertheless, the ratio of pick-up trip length to service trip length becomes rather unfavorable in the smaller operating area, which by far overshadows the empty VKT for repositioning. The pick-up distance is indirectly constrained by the maximum waiting time and therefore similar in the scenarios, while the average user trip lengths are 6.8, 6.1, and 4.3 km for the Default-OA, Small-OA, and Smallest-OA, respectively. Looking at the share of served requests in Figure 3.42c, it turns out that the choice of fleet size per area causes the rate of served requests to have very similar values in the different operating area and fleet size scenarios. Therefore, less vehicles are required per km^2; however, it also has to be considered that the trip density is higher in the central regions with the respective values being 15.6, 18.3, and 25.2 requests / (km^2·h) (from Default-OA to Smallest-OA). When evaluating the share of

served requests as function of the fleet size per request, the result illustrated in Figure 3.42d shows that a smaller number of vehicles per request are required for the same service rate in the smaller operating areas. This is reasonable as trip length and duration are shorter and thereby the number of concurrent travelers is smaller.



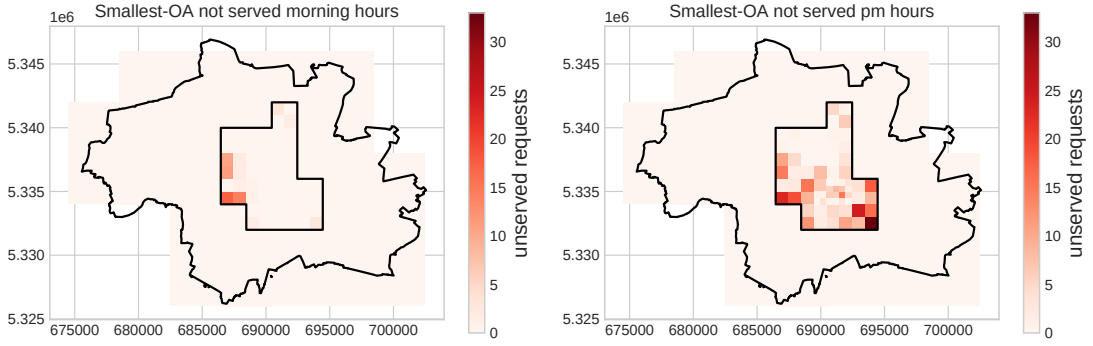
(a) Default operating area with 1750 vehicles in time 03:00 - 12:00

(b) Default operating area with 1750 vehicles in time 12:00 - 24:00



(c) Small operating area with 1274 vehicles in time 03:00 - 12:00

(d) Small operating area with 1274 vehicles in time 12:00 - 24:00

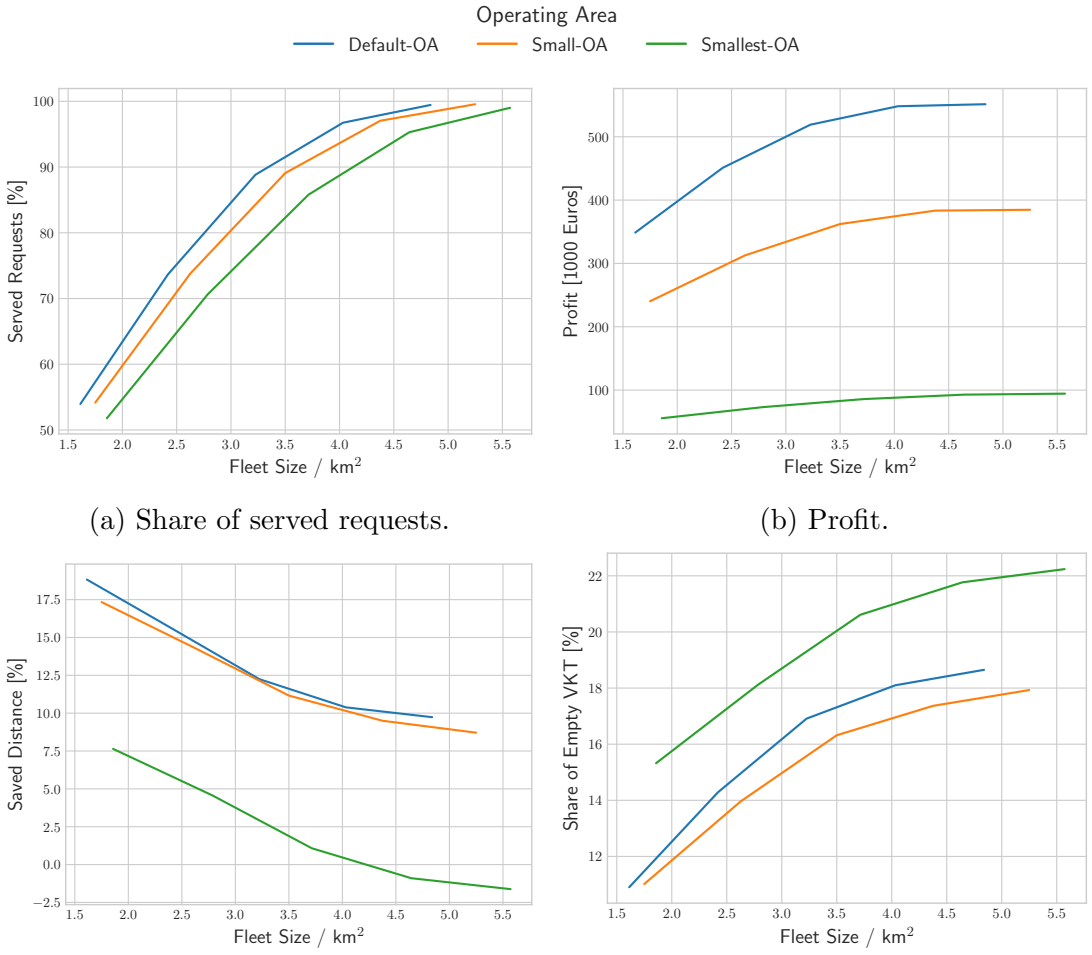


(e) Smallest operating area with 364 vehicles in time 03:00 - 12:00

(f) Smallest operating area with 364 vehicles in time 12:00 - 24:00

Figure 3.43: Total number of unserved requests per forecast zone in the morning and the rest of the day for hailing scenarios. The plots show the area of the city of Munich as reference.

To analyze the unserved requests in more detail, Figure 3.43 displays the spatiotemporal occurrences of these requests based on the zone system and in two time intervals. In most cases, the requests that cannot be served by the AMoD operator are in the outer zones, regardless of the size of the operating areas. The afternoon peak is a little more difficult to handle with the same fleet size as (i) some areas cannot be fully served in this period in the larger operating areas and (ii) the morning peak is no problem in the Smallest-OA scenario. A minor difference between the peaks can be guessed from the plot of concurrent travelers in Figure 3.16c but only through simulation do the effects of vehicle utilization become apparent. Additional, spatial characteristics can be evaluated. For example, the vehicle deficiencies mostly appear in the western part in the morning hours. Fleet operators could re-adjust the imbalance forecasts in these regions. Nevertheless, it has to be noted that the numbers of not served requests are quite small compared to 8000 requests per hour during peak hours in the Default-OA scenarios showing that the repositioning algorithms work well.



(a) Share of served requests.

(b) Profit.

(c) Saved distance.

(d) Empty VKT.

Figure 3.44: KPIs for pooling service for various fleet sizes in different operating areas.

The insights for pooling reflect mostly those from a hailing operation. The curves representing share of served requests per fleet size are again mostly shifted in the x-axis (Figure 3.44a).

As shown in Figure 3.44b, the profit also grows with the size of the operating area. The relative difference between the profit of the Small-OA and Smallest-OA is a little bit higher, which is the result of a higher efficiency difference with respect to saved distance. Due to the RV heuristics and repositioning, the share of empty VKT grows with fleet size. The Small-OA again requires least empty VKT. With $3-4$ %, the difference between the operating areas is smaller than in the hailing case. However, pooling works more efficiently the larger the operating area is, which is why the difference in empty VKT in the Default-OA scenarios is similar to the Small-OA scenarios and the difference to the Smallest-OA scenario grows up to 10 % in distance.

3.8 Conclusion

This section briefly discusses and highlights some of the key results in view of the next chapter, the regulation of AMoD fleets. The change in viewpoint leads to a shift in research questions: this chapter studied how to best serve a given exogenous AMoD demand in a static street network; for the next chapter, AMoD will be analyzed as a part of a transportation system with dynamic networks.

First of all, it has to be noted that the results shown in this chapter are very optimistic because of the network representation (static with free-flow travel times). In congested networks, more vehicles are required to serve the same amount of demand. This can be realized easily by estimating the number of concurrent travelers, which serves as a lower bound for the hailing fleet size. As a vehicle needs more time to serve the assigned request, it becomes available for a subsequent request at a later time. Hence, an operator needs more vehicles at the same time in total.

In general, hailing systems will always increase VKT compared to every person using their private vehicle, i.e., VKT $\geq$ PKT. Pooling is required to reach a state with VKT $\leq$ PKT. Even though it might be less convenient for users, the AMoD pooling service can be designed only to allow minor detours. Considering the time and cost savings from not having to look and pay for a parking lot, pooling should still be a very attractive service design.

With the assumed cost and fare structure, AMoD systems are highly profitable systems. Today's real-world MoD service providers are hardly ever profitable, especially considering the overhead costs for platform operation and staff. However, the profit margins that are possible with AVs seem more than enough to compensate for that, especially considering scaling properties when transitioning to multiple cities.

Operators can benefit from forecasts of demand and supply to improve their profit. In this thesis, several density-based approaches were developed and tested in various scenarios. The performance of the new algorithms surpasses the typical approach with independent zone forecasts. As an alternative for the availability of forecasts, reservation-based systems were studied. However, results showed that either all requests have to include a reservation time or the remaining online requests have a low service rate. Simulations without repositioning or reservation showed that AMoD systems could also serve the same amount of demand without any demand forecasts but are less profitable. Without repositioning, additional fixed costs due to larger fleet sizes overcompensate lower variable costs due to less empty VKT.

The sensitivities with respect to costs are important for the next chapter. With higher fixed

costs, an operator tends to decrease the fleet size, which in turn necessitates more empty VKT to reposition the fleet and pick up passengers. The fleet size is likely also smaller with higher variable costs but to a much lesser degree, and additionally, the operator tries to avoid empty VKT to a higher degree.

For most of this chapter, demand was modeled exogenous and therefore independent of the fare system; only in the dynamic pricing section, users were sensitive to the fares following a logistic function depending on the dynamic pricing factor. When the regulation becomes the topic, the number of AMoD users has to be the result of the competition against other modes of the transportation system, where the fare will also play into the decision-making process. However, the probability of a traveler accepting an AMoD offer will not just depend on the dynamic pricing factor but on more complex relations based on trip characteristics of all available modes. Nevertheless, the service design (general fare) and dynamic pricing will play into the decision-making process.

Unlike for carsharing, where service is often limited to urban centers, the operator is likely to offer its service in the whole urban area. With the assumed fare and cost structure, the AMoD operator can make more profit with larger operating areas. Of course, this statement is made with assumed exogenous demand. However, the assumed price level can be expected to attract a significant amount of demand. This is a positive sign as PT services are typically developed best within the city centers and could be complemented with connections from, to, and between the outer regions.

An attractive AMoD service with a large user base has the potential to save a lot of urban space. In the 5 % demand scenarios, less than $2,000$ vehicles served more than $100,000$ trips. This amount is by far more than the average number of trips a private car makes in a single day. Even though AMoD vehicles will still require some space for pick-up and drop-off, the amount of urban space that can be saved is a great outlook. Actually, this chapter already contained a policy study considering the question of whether AMoD vehicles should be allowed to park on-street or have to return to depots. From an operator's perspective, street parking is clearly beneficial. City administrations will have to weigh the value of parking space against the additional empty VKT that is necessarily produced by banning AMoD street parking. Moreover, this study assumed the case where operators made return trips to depots instead of circling around their AVs, which will be difficult to prohibit. A return duty, as stated in the current German law, is counterproductive in the context of AMoD and should be abolished.

Finally, it has to be noted that the computational complexity of AMoD operations is quite significant, especially for pooling services. For a real-world operator, it makes sense to invest the required computational resources to gain additional profit as a few thousand € per day and per city accumulate to a significant amount over time and multiple operating areas. Contrarily, from a transportation systems perspective, the use of heuristics to speed up the simulation of operations is likely sufficient in most cases. Nevertheless, the most relevant service design parameters — fleet size and fares — affect the service rate significantly and should be considered in transportation system models.

Chapter 4

Regulation of AMoD Systems

This chapter studies the impacts of AMoD systems on the transportation network and hhow a regulating public entity can steer the effects. This public entity can be on municipal, state, or federal level and is denoted by the regulator in this chapter.

In most transportation system models, the system state is given by some form of equilibrium of supply and demand. The transportation network determines the supply side, and travelers choosing their mobility behavior represent the demand side. An equilibrium is typically established because of feedback loops. On the one hand, travelers choose a particular route the better the travel time is on this route; on the other hand, the travel time on this route becomes worse as more travelers choose it and it becomes congested.

Similarly, an equilibrium can be expected in the presence of an AMoD service. Demand is attracted until the fleet is highly utilized and cannot offer a high service quality for new requests. Moreover, the addition of more vehicles generates additional system capacity but also creates extra costs.

When measures for the current transportation systems should be developed by transportation planners of or for public entities, bi-level modeling is often the approach of choice. On the upper level, the measures are adapted into the transportation system and optimized based on the performance of the lower level transportation model, in which travelers optimize their behavior based on the adaptions.

A new tri-level framework was developed, including the AMoD service operator as an additional decision-making player next to the regulator and the travelers. A separation of (time) scales is applied once again to formulate the problem. Public entities have several possibilities to regulate AMoD directly, which can range from prohibiting service altogether, some service design aspects (e.g., no hailing, or no service in certain areas), to changing the cost structure by enforcing a road toll, parking fees, licenses, or even handing out subsidies (e.g., for first/last mile service). Moreover, other regulatory push and pull measures, like making private vehicles more expensive or increasing the public transport budget, affect the AMoD system indirectly because they can modify the mobility behavior of travelers. Unlike previous studies investigating single regulatory scenarios, in which operators did not adapt their service, the newly developed framework considers operator adaptations and offers the possibility to optimize certain numeric variables, e.g., the height of a toll.

Parts of this chapter were already published in Transportation Research Part C before this thesis [Dandl, Engelhardt, et al., 2021]. Besides minor editorial changes and adaptions of notation, this chapter consists of direct and indirect citations of this reference. For reasons of readability, it is refrained from repeating the citation after every paragraph.

In the following sections, a mathematical tri-level problem formulation and a solution approach are introduced to tackle the regulation of transportation systems in the presence of an AMoD service. The new framework is general, and for a case study, the regulatory measures and fleet operational variables have to be determined and their impacts modeled. This step is conducted with the help of an agent-based simulation model applied to a case study based on the same area as in the previous chapter, i.e., Munich, Germany. Finally, this chapter concludes with a more general discussion.

4.1 Problem Description

This chapter models a future transportation system with three key actors — the regulator, the AMoD service planner, and travelers/system users. The regulator sets the political, legal, and organizational framework; the AMoD service planner offers a profit-maximizing service depending on the existing regulatory environment; and each traveler chooses the best of her available mobility options. On the one hand, the choices of travelers affect the utilization of the road network and PT system as well as congestion levels, environmental impacts, and the profit/loss of both the AMoD business and the public agency. On the other hand, regulators and AMoD service planners can influence the decisions of travelers. The regulator has direct and indirect methods to change travelers' mode decisions and to impact the AMoD service planner's decisions. Increasing parking fees makes PVs less attractive [Washbrook et al., 2006]. Subsidies to PT or AMoD services represent indirect methods as these financial policies could motivate AMoD service planners to offer cheaper trips. A road toll is an example that affects users both directly and indirectly. PV users directly pay the toll and AMoD companies – if not exempted – are likely to pass the additional costs on to their passengers. Furthermore, a zone-based distance-dependent toll can influence route choice decisions of AMoD operators and individual drivers.

This thesis denotes (i) the decision variables of the regulator by α^y, where y is the index for different regulations; (ii) the decision variables of the AMoD service planner by β^s, where s is the index for different service design parameters; and (iii) the mode-choice decision variable of traveler r for mode m by m_r ($m_r = 1$ for the chosen mode, 0 for the others). For the sake of simplicity and clarity, this thesis limits the set of mobility options to PV, one AMoD service and PT with walking access and egress.

The decision-making processes of the three different actor classes represent a non-cooperative game. The study assumes there is a hierarchy of decision making: regulators effectively set the rules of the transportation game, AMoD service planners define their service, and travelers make their mode decisions based on the regulator rules and the AMoD planner's service. This approach reflects reality as policies and regulation changes generally require a long planning period; whereas, an AMoD service planner can likely change/adapt their service characteristics relatively frequently (e.g., monthly or quarterly). Finally, a subset of travelers can update modal decisions on a weekly, if not daily, basis. This approach represents a separation of time scales.

The goal of the tri-level decision problem is to find a set of regulator decisions $\alpha^{y,*}$ that maximizes social welfare W subject to the conditions that (i) the AMoD business optimizes its service design to maximize profit P for a given set of regulations α^y and (ii) travelers

optimize their travel behavior (combined mode and route choice) according to their personal utility $U_r = \sum_m m_r \cdot U_r^m(\alpha^y, \beta^s)$, subject to regulations α^y and the AMoD service design β^s.

$$\alpha^{y,*} = \arg\max_{\alpha^y} W\left(\alpha^y, \beta^{s,*}(x^r), \{m_r^*(\alpha^y, \beta^{s,*})\}_r\right) \tag{4.1}$$

$$\beta^{s,*} = \arg\max_{\beta^s} P\left(\beta^s, \{m_r^*(\beta^s)\}_r | \alpha^y\right) \tag{4.2}$$

$$m_r^* = \arg\max_{m_r} U_r\left(m_r | \alpha^y, \beta^s\right) \qquad \forall r \tag{4.3}$$

Here, the | symbol denotes that the right-hand variables are set by a higher-hierarchy player.

4.2 General Solution Approach: Multi-Level Bayesian Optimization

Equations (4.1)-(4.3) represent a tri-level mathematical program. To solve the tri-level math programming problem, this thesis employs a simulation-optimization solution approach. The simulation-optimization approach includes the transportation model at the lowest level to capture the behavior of travelers given the middle and upper level decisions. The travelers are individually optimizing their own utility and the agent-based simulation solves for user-equilibrium. Additionally, the simulation-optimization solution approach employs Bayesian optimization (BO) to optimize the AMoD service planner's decisions in the middle level and to optimize the regulator's decisions at the highest level.

The transportation model performs all traveler optimization processes and returns the aggregated quantities for social welfare W and AMoD service planner profit P as functions of α^y and β^s, which are inputs to the transportation model. Therefore, the transportation model must be sensitive to the regulations α^y and the AMoD service design β^s. Section 4.3 describes the details of the transportation simulation model and model components.

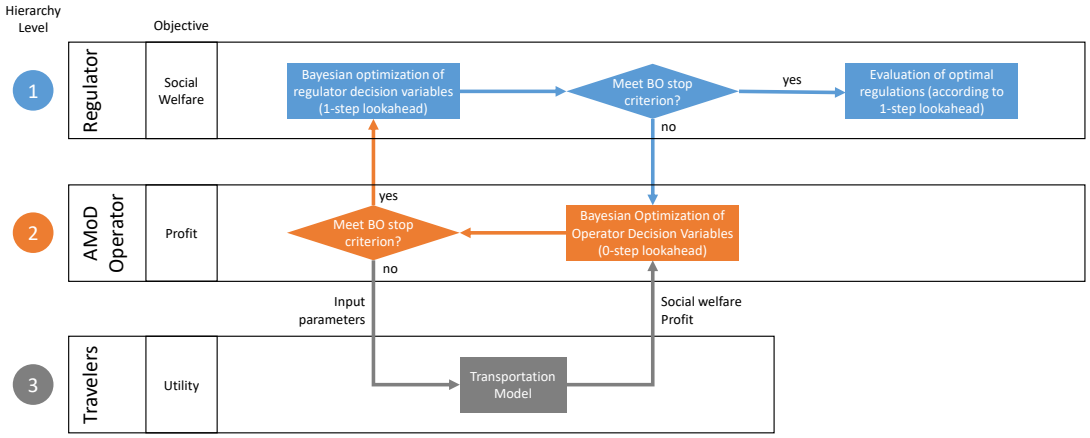


Figure 4.1: Flowchart describing the processes of the solution approach on a high level.

The two-level optimization process is a two loop process. The AMoD service planner optimization loop iterates until a solution to equation (4.2) is found. Then the process jumps to

the next iteration of the regulator level. This two-level process continues until the upper-level solution converges. Fig. 4.1 displays a flowchart describing the procedure for an arbitrary transportation model that returns W and P given α^y and β^s.

A naive approach would re-start the search for $\beta^{s,*}(\alpha^y)$ from scratch for every hyper-plane in the solution space defined by a new set of regulator variables α^y. Conversely, BO can help reduce the number of lower level iterations and thereby overall computational effort.

Both regulator and AMoD service planner level optimization can be non-convex (multi-dimensional) problems. In these types of problems, an optimization algorithm needs to incorporate gradient search techniques to move closer to optima, but avoid getting stuck in local optima. Since a single function evaluation means running a full transportation model simulation, the number of explicit evaluations should be as small as possible. BO in general is an iterative process for multi-dimensional non-convex optimization problems with two components: a *surrogate model* representing the objective function and an *acquisition function* determining the next point in the solution space for which the original function, in this case the transportation model, should be evaluated.

The task of the surrogate model is to infer knowledge from prior iterations; the task of the acquisition function is to exploit this knowledge to search areas where good objective function values can be expected while also exploring the solution space to avoid ending up in a local optimum.

Additional to this nice general feature, BO provides another key advantage in the regulator-AMoD service planner setting: the surrogate function can infer approximations for the profit-optimal solution $\beta^{s,*}(\alpha^y)$ from prior iterations in hyper-planes with different regulatory setting $\alpha^{y,'}$. The surrogate function develops a sense of closeness in both α^y and β^s directions with regard to the underlying profit function $P(\alpha^y, \beta^s)$. The 'closer' another setting $(\alpha^{r,'}, \beta^{s,'})$ is, the larger its effect on the inference of unknown data points. This exploitation of prior knowledge reduces the number of required iterations.

In this framework, Gaussian Processes are used to model surrogate functions for AMoD profit $P(\alpha^y, \beta^s)$ and social welfare. In order to anticipate an AMoD operator's reaction, the surrogate function for social welfare actually approximates $W(\alpha^y, \beta^{s,*}(\alpha^y))$. It considers only the social welfare data points from profit-optimized variables in the solution space $(\alpha^y, \beta^{s,*}(\alpha^y))$. Hence, this social welfare surrogate function is defined on the space of the regulator variables.

In the following two paragraphs (Gaussian Process and Acquisition function), single-level BO with Gaussian processes is briefly described. For a clear presentation, typical mathematical notation without the transportation context is utilized.

Gaussian Process: Surrogate functions should be smooth representations of the true underlying function. They are estimated from prior data points. Like many other BO models, this thesis applies a Gaussian Process to infer the surrogate function value of any point in the solution space. Additionally, Gaussian Processes provide estimates of the uncertainty associated with an estimation. This type of surrogate function is described through normal distributions $\mathcal{N}(\mu, \sigma)$ at each point of the solution space.

Let $\tilde{X}$ be a vector space (solution space), $f : \tilde{X} \to \mathbb{R}$ be an underlying function, and $(x_p, f_p = f(x_p))$ with $x_p \in \tilde{X}, f_p \in \mathbb{R}$ denote data points from prior explicit evaluations of f.

Then a Gaussian process denotes a surrogate function

$$f^S(x) \sim \mathcal{N}(\mu(x), \sigma(x)) \tag{4.4}$$

Both μ and σ are inferred from the set of prior data points (x_p, f_p) with the help of covariances. Let $X = \cup^p x_p$ be the set of prior solution space data points, $x, x_1, x_2 \in \tilde{X}$ be points in the solution space, and $k(x_1, x_2)$ be the covariance of two data points in the solution space. Together with equation (4.4), the following equations describe a Gaussian process:

$$\mu(x) = \sum_{p \in X} \sum_{q \in X} k_p(x) \mathbf{K}_{pq}^{-1} f_q \tag{4.5}$$

$$\sigma^2(x) = k(x, x) - \sum_{p \in X} \sum_{q \in X} k_p(x) \mathbf{K}_{pq}^{-1} k_q(x) \tag{4.6}$$

where $\mathbf{K}$ denotes the constant covariance matrix, where each entry $\mathbf{K}_{pq} = k(x_p, x_q)$ for all prior data points x_p, x_q. Furthermore, $k_p(x) = k(x_p, x)$ describes the covariance between a prior data point and the point in the solution space x, for which the Gaussian Process should be computed. $\mu(x)$ can be used as surrogate function and $\sigma(x)$ gives information about the uncertainty of the approximation. This thesis employs the commonly chosen Matern kernel:

$$k(x_1, x_2) = \frac{1}{2^{\zeta-1}\Gamma(\zeta)} \left(2\sqrt{\zeta}\,\|x_1 - x_2\|\right)^{\zeta} \mathcal{H}_{\zeta}\left(2\sqrt{\zeta}\,\|x_1 - x_2\|\right) \tag{4.7}$$

where Γ is the Gamma function and $\mathcal{H}$ the Bessel functions of order ζ and $\zeta = 5/2$ is set as in [Y. Liu, Bansal, et al., 2019]. Essentially, the covariances determine the contribution of each prior observations $f(x_p)$ on the approximation of f for an arbitrary point x in the domain.

Acquisition Function: The iterative nature of the BO is controlled by the acquisition function. In the n-th iteration, the point $x_{n+1} \in X$, for which the function f should be evaluated explicitly, is determined by the maximization problem:

$$x_{n+1} = \arg\max_x A_n(x) \tag{4.8}$$

where $A_n : \tilde{X} \to \mathbb{R}$ is the nth-step acquisition function. This function in general has the same dimension as the data points and (in the best case) the unknown underlying function f. However, with the help of $\mu(x)$ (representing the surrogate function) and $\sigma(x)$ from the Gaussian Process, this optimization is performed on an analytical, smooth and differentiable function defined over the whole domain with cheap function evaluations.

The surrogate function is an approximation of the underlying function f. If the surrogate function would be chosen as the acquisition function, the iterative process is likely to end up in a local rather than the global minimum. It might be possible that the approximation further away from the data points is rather bad and even though the surrogate function decreases, the underlying function becomes better with respect to the overall objective. To avoid this phenomenon, the acquisition function generally contains a measure of the uncertainty as well.

Various possible definitions exist in literature, see e.g. [Y. Liu, Bansal, et al., 2019]). The Upper Confidence Bound method is applied here:

$$A_n(x) = \mu_n(x) + \kappa_n \sigma_n(x) \tag{4.9}$$

$$\kappa_n = \sqrt{2 \log \left(\frac{\pi^2 n^{2+\dim/2}}{3\delta} \right)} \tag{4.10}$$

where dim is the dimension of the solution space and $\delta = 0.1$ is chosen as in [Srinivas et al., 2012]. The inclusion of n in the definition of κ_n helps to adapt the weight of the uncertainty for higher iterations. However, it also causes the algorithm to keep searching in regions with low expectation value in case there are a larger number of data points. In order to exploit prior knowledge to a higher degree, κ_n was limited to 1 for the later stage of the optimization process in the case study.

Parallel Initialization of Two-Level BO Process: The previous explanation described the application of BO on typical one-level optimization. In this novel framework, BO is used for a two-level optimization. The regulator optimization represents an ordinary BO of social Welfare W with respect to the variables α^y. However, the AMoD service planner optimization is only applied in a hyper-plane, namely the AMoD service design variables for a given set of regulatory variables. Hence, the optimization only takes place along β^s, but data points outside α^y hyper-plane of the current upper level iteration can help infer the surrogate function for the optimization of $P(\alpha^y, \beta^s)$.

In general, there is no strict rule how to set the initial set of data points in BO. Obviously, the choice affects the path of the iterations, but due to the uncertainty being part of the acquisition function, the exploration of the solution space should in most cases be achieved. However, a good initial distribution of data points can help the procedure.

In this thesis, the initialization process is divided in two stages: first, the operator variables β^s are optimized for the case of no regulations. An initial set of simulations are generated by Sobol sequences in the operator solution space. This low-discrepancy sequence method creates a good spread of data points in high-dimensional spaces [Sobol, 1976]. If the solution space is constrained, the corners of the variable space are added to the initial set of simulations as these would likely be visited in the next iterations. Then, the BO method determines the profit-optimal solution in the operator space in the no regulation case. After that, an initial set of data points in the regulator space are created by Sobol sequences (with the corners of the regulator space for constrained variables). By initializing processes for a multitude of regulatory settings $\{\alpha^y\}$ in parallel and joining the information (data points) after each evaluation, each operator optimization process benefits from the results of 'close' regulator-variable hyper-planes, which can reduce the number of necessary iterations even further.

4.3 Model

The solution approach described in the previous section is general, and valid with any transportation model. However, this thesis proposes an agent-based transportation system simulation model at the lowest level, in order to model the experiences of individual travelers

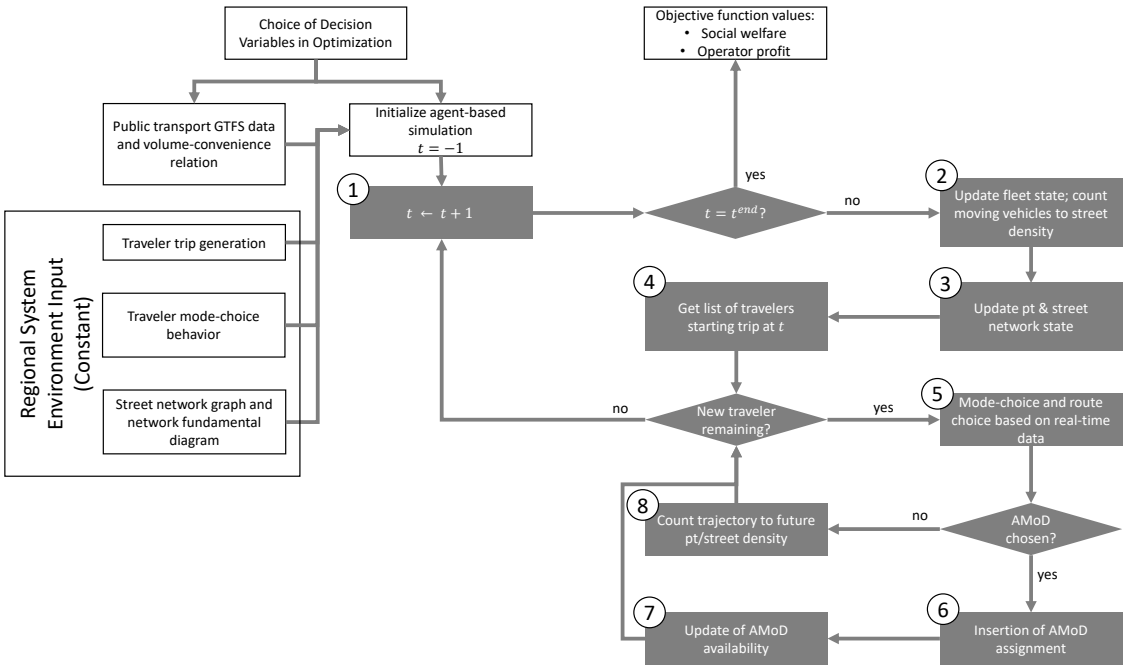


Figure 4.2: Flowchart describing the processes of the transportation model.

(particularly, AMoD users); the impacts of AMoD service decisions and operational strategies; and the impacts of specific regulation, in the appropriate detail to capture the experiences of travelers and the system impacts. This agent-based transportation model can evaluate social welfare W and profit P for a certain set of regulatory and AMoD parameters α^y and β^s.

The transportation model executes a loop (with index i) over time with time steps of size Δt ($t \leftarrow t + \Delta t$); the time-step is one minute in the following case study. Figure 4.2 illustrates the model inputs and process flow at a high level. The AMoD aspect is nearly identical to the simulation framework described in Section 3.6.1, but the model contains a lot of additional components to represent a complete transportation system.

After incrementing the simulation time (1), the state of AMoD vehicles is updated (2). During this step, the framework counts how many fleet vehicles are currently moving on the streets and add them to the density of the street network. Together with the density of PV and PT travelers, these determine the update of travel times and crowding (3). For new travelers want to start their trip (4), a combined mode- and route-choice model, which is based on real-time information will be applied (5). From an AMoD operator view, this includes the user-operator interaction with request, offer and user decision, just as in the simulation model of the previous chapter. In the following case study, travelers can choose between PV, AMoD, and PT[1]. If AMoD was selected, the AMoD operator inserts the pick-up and drop-off stops into its vehicle routes (6) and updates its availability for new requests (7), i.e., guaranteeing the trip of the just assigned request is already considered for the next request. Thereby, the fleet will not create good offers anymore when it is at its capacity. Finally, the travelers

[1]Pedestrians and bicyclists with shorter trips are assumed to stick to their travel behavior and are excluded from this case study.

choosing PV or PT are counted to the density of future time steps according to their projected trajectories (8). The following subsections describe each of the components in more detail.

Computation time is one of the key challenges of this tri-level optimization problem as the upper levels require multiple runs of the transportation model. Hence, the subsequently described model aggregates where possible while keeping the agent-based nature of travelers and AMoD vehicles. Moreover, cities are usually not spatially homogeneous regarding the density of inhabitants, work places, PT infrastructure, etc. To address the trade-off between computational efficiency and modeling accuracy, the proposed model separates the study area into two zones Z: an inner city zone z_I and an outer city zone z_O. Note that the model is not limited to two zones and could be extended to more zones.

The transportation model must be sensitive to the upper level decision variable inputs. However, the model can remain relatively aggregate for portions of the transportation system that are not significantly impacted by potential AMoD regulations. Hence, the selection of decision variables influences the types and resolution of the transportation model components. Therefore, the next step is to define the decision variables that will be analyzed in the case study to follow.

4.3.1 Selecting a Set of Decision Variables

There are many policies a city can test to improve the performance (i.e. social welfare) of a transportation system. Often, these policies are usually studied independently of each other, thereby ignoring potential synergistic and opposing effects among multiple policies. This chapter and especially the general problem (Section 4.1) and high-level solution approach (Section 4.2) aim to enable the optimization and analysis of a combination of policies. Nevertheless, the study focuses on a selected set of policies and AMoD service planner responses (i.e. decision variables) to keep the system computationally tractable.

Internalization of External Costs of Private Vehicles: Even though PVs wear the streets, cause congestion, emit harmful pollutants, and require valuable space for parking in urban areas, PV owners do not pay (i.e. internalize) these cost in most cities. Higher costs for parking spaces and a toll for driving on urban roads could decrease the attractiveness of using the PV. AMoD services also utilize the street network and are not exempt from parking fees and road tolls in general. However, as they are likely to spend much less time parking [Fagnant and Kockelman, 2014], this case study assumes that they do not need to pay parking fees. Moreover, this choice seems reasonable in the context of the previous chapter. Toll costs and parking fees are likely treated as variable and fixed costs by the operator, respectively. Hence, parking fees on AMoD would tend to decrease fleet size and increase VKT.

PT Budget: If the regulator wants to increase demand for PT, they should improve PT service quality. In most US cities to improve service quality, the regulator would need to provide a larger budget to the PT agency, who would subsequently determine how to improve their transit service. Increasing frequencies is necessary to accommodate more travelers and improve their comfort as the PT vehicles are less crowded. The costs for increasing frequencies can serve as an input. However, it is not clear how many people will choose PT as their mode;

the revenue and therefore the required PT budget are an output of a simulation run. Hence, an iterative process would be necessary with PT budget as regulator variable. For simplicity, the regulator can directly influence the PT frequencies in this model.

Fleet Size Limitation: City administrations could enforce an upper bound on the AMoD fleet size. This could be achieved by licensing of AMoD vehicles. Only licensed AMoD vehicles would be allowed to serve customers and the regulating authority controls the number of licenses that are handed to the AMoD provider.

AMoD Service Planner Decision Variables: The operation of AMoD services can vary in many details. The most influential service design parameter is fleet size. As shown in the previous chapter, it determines the maximum number of travelers that can be served at one time and is one of the main factors for vehicle availability and traveler wait times. If a city enforces a fleet size limitation, the upper bound of this variable is set by the regulator. This case study also looks into the price structure of an AMoD service planner as this influences the demand and thereby indirectly the vehicle movements and fleet distribution.

Since substantial improvements to the transportation system by a pure ride-hailing service without shared rides are unlikely, this thesis assumes that the regulator constrains the AMoD operator to provide a pooling service. AMoD customers may be driven alone, but will share their ride if a match is possible.

4.3.2 Traveler Mode Choice Model

A traveler r is modeled as an agent that wants to travel between an origin-destination pair od at a time τ_r. Both start location x_r^p and end location x_r^d are within the boundaries of a pre-defined study area. For mode choice, the traveler considers the set M of following travel options: PV, PT (with walking access/egress), and AMoD.

As stated in the problem formulation, each agent r tries to maximize the utility related to a trip. In order to represent stochastic differences in travel behavior, mode-choice is determined by a logit model. The probability of traveler r choosing mode m is given by:

$$P_r(m) = \frac{\exp[U_r^m]}{\sum_m \exp[U_r^m]} \tag{4.11}$$

where U_m are the utilities of the modes. This case study models the utility of each mode as the negative of the mode's traveling costs, i.e., $U_r^m = -C_r^m$. For the sake of simplicity, the case study uses one value of time c^{VOT} for all time-related factors (e.g., in-vehicle time, wait time, walk time, etc.) in the utility/cost function.

For PV trips, the fastest route is considered. The travel time t_{od}^{PV} is weighted by value of time c^{VOT}, the travel distance d_{od}^{PV} is weighted by the distance-dependent vehicle costs $c^{D,PV}$ and the utility can also be impacted by the toll $C_{od}^{toll}(\tau_r)$, and parking fees $C_{od}^{park,PV}(\tau_r)$ for OD-relation od of traveler r at time τ_r. The explicit toll and parking fees depend on the decisions of the regulators and are described in sections 4.3.5 and 4.3.5, respectively. Additionally, the model includes an intercept term c^{PV} to calibrate the modal split in the model in the absence

of AMoD service and to reflect unobserved attributes representing the attitude towards PV compared to the other modes.

$$C_r^{PV} = c^{VOT} \cdot t_{od}^{d,PV} + c^{D,PV} \cdot d_{od} + C_{od}^{toll}(\tau_r) + C_{od}^{park,PV}(\tau_r) + c^{PV} \tag{4.12}$$

For PT, the model considers the fare f_{od}^{PT}, the access and egress distances d_r^{walk} to and from PT stops with a walking speed v^{walk}, the number of transfers N_T, the transfer penalty c^T, crowding of the PT system η^{PT}, and the travel time t_{od}^{PT}. Each transfer represents a risk of additional travel time. Delays on one line can cause a traveler to miss the next connection; the resulting additional time for the traveler depends on the frequencies of the PT lines. Furthermore, crowding impacts utility through the value of travel time with the help of a monotonously increasing function $g^{VOT}(\eta^{PT})$. In other words, the perceived travel time increases with the level of crowding in vehicles. This is modeled as a piece-wise linear function g^{VOT} with supporting points $(\eta^{PT}, g^{VOT}(\eta^{PT}))$ described in C.

$$C_r^{PT} = f_{od}^{PT} + g^{VOT}(\eta^{PT}) \cdot c^{VOT} \cdot t_{od}^{d,PT} + c^{VOT} \cdot v^{walk} \cdot d_r^{walk} + c^{VOT} \cdot t_r^{w,PT} + c^T \cdot N_T \tag{4.13}$$

For AMoD, the model assumes that the fare, the travel time, and the wait time affect the utility associated with AMoD and thus the mode choice of travelers. The AMoD service price/fare is one of the AMoD service planner's decision variables and will be described in more detail in section 4.3.5.

$$C_r^{AMOD} = f_r^{AMOD} + c^{VOT} \cdot t_r^{d,AMOD} + c^{VOT} \cdot t_r^{w,AMOD} \tag{4.14}$$

Modeling Assumptions: As this thesis aims to estimate the long-term effects rather than the short-term effect that AMoD can have on PV ownership and the transportation system, travelers do not have a PV ownership attribute. Instead, all travelers are given the choice to use PV, but at full costs (i.e., $c^{D,PV}$ includes depreciation, insurance, maintenance, fuel, etc.) instead of just operating costs.

For the sake of simplicity, the model does not incorporate impacts of dynamic costs (toll, AMoD pricing) on the departure time choice of travelers.

Another modeling assumption typical to large-scale transportation models is that travelers make their decisions immediately. Therefore, just as in the previous chapter about the operation of AMoD systems, the operator will know immediately, whether a request accepts an offer or chooses an alternative mode and rejects the offer.

4.3.3 Network Model

Route choices affect the costs of all modes. As modeling route choices is a computationally expensive process, the proposed framework pre-processes both the PT and the street network and makes some simplifying assumptions that enable computationally efficient network modeling. PT stations are matched and connected to the closest street network node with zero cost. Traveler origin and destination locations are matched to the nearest street network access/exit nodes.

Street Network and Routing Model: PVs and AMoD fleet vehicles drive in the street network graph $G = (N, E)$. The attractiveness/utility of these two modes is severely influenced by route travel times in the network. The travel times, in turn, depend on the state of the network, which is affected by the number of vehicles on the streets. In order to be useful in the larger modeling framework, the street network model needs to reflect traffic dynamics. Furthermore, spatial granularity is necessary to model the spatio-temporal availability of AMoD vehicles and pooling of travelers for shared rides. Modeling AMoD in a microscopic traffic simulation [DANDL, BRACHER, et al., 2017] would satisfy both criteria, but is computationally very demanding and not suitable for the upper level optimizations.

This chapter of the thesis employs an approach that combines spatially granular but computationally efficient routing with macroscopic traffic dynamics. The approach involves preprocessing all vehicle paths based on free-flow velocities. Therefore, the efficient algorithm described in Appendix B.2 can be used to create routes.

The travel times on each path are scaled using factors ψ_c^t derived from separate (simplified) network fundamental diagram (NFD)s for the outer and inner city. Hence, vehicles always travel on the path that has the shortest travel time under free-flow conditions. However, the edge (i.e. link) travel times in the transportation model simulation are scaled according to the traffic state on each edge. The travel time on edge e in cluster/zone $z \in Z$ at simulation time t is given by

$$t_e^t = \psi_z^t t_e \tag{4.15}$$

where t_e is the free-flow travel time on this edge. The cluster travel time factors ψ_z^t are updated each time step t based on the vehicle density in the respective part of the street network. The density comprises the number of PVs and AMoD vehicles driving at time t, which transport the agents within the study area, as well as the vehicles starting or ending their trips outside of the study area, which only act as background traffic. The number of background traffic vehicles needs to be calibrated. The flow q_z^t and thereby the macroscopic velocity and travel time factor in cluster z at time t is derived from the NFD $q_z = q_z(k_z)$ and $\bar{v}_z = q_z/k_z$.

As in [DANDL, TILG, et al., 2020], the travel time factors are given by the functional form

$$\psi_z^t \sim v_{1,z} \left(\frac{1}{\bar{v}_z} + \frac{1}{v_{2,z}} \right) \tag{4.16}$$

and the parameters $v_{1,z}$ and $v_{2,z}$ are fitted according to a travel time comparison with a dynamic traffic assignment simulation (in a pre-processing step). In order to reduce oscillations, the moving average for k_z of the last 5 time steps is used to update the travel time factors.

In order to reduce the computational burden, PVs are not rerouted and tracked in the model. Only AMoD vehicle states are updated each time step because they affect the offer the AMoD operator can make to new travelers and therefore their decisions.

This NFD-based approach to scaling network travel times precludes meaningful evaluation of route choices. However, this approach effectively approximates the quantity affecting traveler's mode-choice and AMoD vehicle-passenger matching: the travel time between two points.

PT Network Model: A PT system includes the spatial network layout of all lines, the schedule for each each line, as well as the specification of vehicle types (e.g., buses and trams).

Such PT systems are specified in the widely used GTFS format[2]. GTFS specifies the PT system including schedule, stops, routes, and vehicles. To process routing queries in such a system, OpenTripPlanner[3], which is based on OpenStreetMap[4] is a useful tool. By specifying origin and destination, as well as a departure time, possible PT trips are found. Each trip includes access, egress, and transfer walking distances, total travel time and number of transfers. These parameters are required inputs for the traveler choice model.

This case study aims to investigate a large operating area with a high number of travelers and thus computational time is of essence. In order to increase efficiency, routing queries in the PT system between all stop-to-stop combinations are pre-processed. The result is a data table (i.e. skim matrix) in which the walking distance, travel time, and number of transfers for each stop-to-stop combination is stored. This procedure allows to reduce the OpenTripPlanner query to a simple look-up in the pre-processed data thereby substantially decreasing the computational cost of such queries. As the waiting time at the station depends on the schedule and whether the PT service operates on time, it is assumed that the traveler arrives just in time, i.e., with $t_{wait}^{PT} = 0$.

Another PT attribute affecting a traveler's mode choice given by equation (4.13) is the crowding within PT vehicles. This is modeled in an aggregated fashion similar to the street network. Whenever a traveler chooses PT, the system-wide count of PT travelers increases by one for every time step – until the arrival. Crowding represents the equivalent of congestion of road-bound traffic. The literature suggests that the perceived travel time of (standing) passengers increases mostly linear with the number of passengers per vehicle. Therefore, the effects of crowding on the travel time can be represented as a (value of) travel time factor. Studies for various large cities suggest a maximum factor of 2.0 [Tirachini, Hurtubia, et al., 2017]. Due to the lack of corresponding studies for the city of Munich, the average maximum factor to model the effects of travel time is applied in this case study. The crowding factor η_t^{PT} is defined as the ratio between the number of PT travelers n_t^{PT} and the total PT capacity Ω_T^{PT}.

$$\eta_t^{PT} = \frac{n_t^{PT}}{\rho_T^{Pt}} \tag{4.17}$$

$$\rho_T^{PT} = \frac{1}{T} \sum_l \sum_{\zeta_l^T} \rho_l^{PT} = \sum_l \frac{|\zeta_l^T|}{T} \cdot \rho_l^{PT} = \sum_l \nu_l^T \cdot \rho_l^{PT} \tag{4.18}$$

where ρ_l^{PT} is the capacity of a PT vehicle on line l (train, tram or bus) and $|\zeta_l^T|$ is the number of trips on this line during a time interval T (chosen to be one hour) and ν_l^T the derived frequency. Similar to the street network, PT users traveling from or to areas outside of the study area need to be included in a calibration process.

For the estimation of social welfare, operating costs and emissions have to be evaluated. Therefore, operating cost C_T^{PT} and emissions E_T^{PT} of the PT system during time interval T

[2] https://developers.google.com/transit
[3] https://www.opentripplanner.org/
[4] https://www.openstreetmap.org

are approximated based on the frequencies:

$$C_T^{PT} = T \sum_l \nu_l^T \cdot d_l \cdot c_l \tag{4.19}$$

$$E_T^{PT} = T \sum_l \nu_l^T \cdot d_l \cdot w_l \cdot e_w \tag{4.20}$$

where d_l, c_l and $w_l \cdot e_w$ are the mean length, cost, and emissions of a PT vehicle trip (from start to end station) of line l, respectively. The assumed cost values reflect typical per schedule-km costs beyond operating costs for the respective vehicle type (bus, tram, train) on line l. Together with the revenue of PT travelers, they represent the PT budget.

4.3.4 AMoD Fleet Control Model

AMoD vehicles are explicitly represented as agents in the transportation model, which are controlled centrally by the AMoD operator. The control model follows the same process flow as described in the previous chapter: a traveler r requests information about a possible trip, the AMoD operator generates an offer containing fare, expected waiting and driving time based on the current state of its system and possible insertions of r into the existing vehicle plans. Then, the traveler chooses whether to use AMoD or another mode, which translates to a rejection of the AMoD offer. The vehicle plan is updated in case of an acceptance and reset in case of a rejection.

As mentioned in the conclusion of the last chapter, real-world operators are likely to optimize all decision levels. However, here the operational decisions are based on the IRS strategy without vehicle plan re-optimization and with RV heuristics described in Section 3.3.2. Even though optimization potential is not exhausted, the results of the last chapter show that the IRS strategy can achieve decent solutions with RV heuristics. Most importantly, the required computational effort for re-optimization of pooling assignments or for considering every vehicle for very large fleets would severely slow down the general framework.

The fleet operator still assigns maximum wait time (τ_r^w) and relative detour time (τ_r^δ) constraints for a request r. These no longer determine whether the request will accept or reject an offer but are still useful tools to limit the vehicle search space (τ_r^w) and possible insertions (τ_r^w and τ_r^δ). These level-of-service parameters define the service design and, thereby, the user experience of the envisioned service and are chosen as in the case study of the last chapter. However, since travel times are no longer static, it is possible that constraints, which are valid for a request r at time τ_r, become infeasible over time. Since the constraints do not actually determine traveler decisions, this (real-world effect) is not considered as a problem. However, the addition of new requests is prohibited unless all time constraints can be satisfied to limit delays.

The control objective is chosen just as in equation (3.16) of the previous chapter:

$$F_{\xi_v}^S = c^D \cdot d_v^\xi + c^{VOT} \sum_{r \in R_v^\xi} \left(t_r^w + t_r^d \right) \tag{4.21}$$

With respect to the mid-term strategies, dynamic pricing is applied with the utilization based model. Moreover, this case study refrains from using a repositioning strategy. First of

all, forecasts of future demand and supply are not available here because demand is endogenous and cannot be derived from historical data. As shown in the results of the previous chapter, this will necessitate larger AMoD fleets to reach a similar level of service but with less empty VKT. This might be aligning better with the regulator objective depending on the desired trade-off between required space on and off the road.

4.3.5 Modeling the Impacts of Decision Variables

The transportation model has to be sensitive to the decision variables stated in Section 4.3.1. The following describes how the decision variables interact with the other parts of the transportation model.

Parking Fee Model: Parking fees for PVs are generally dependent on the location and the amount of time a vehicle is parked. In an activity and agent-based travel demand model, the parking time can be approximated very well by the time of a given activity. For trip-based models, the parking costs become much more difficult to approximate. Furthermore, there is no way of keeping track of vehicle consistency, i.e., a traveler who chooses a PV to get to work is very likely to choose the car back on the way home. Unfortunately, only trip-based data is available in the investigated case study (see section 4.4). For the sake of simplicity, it is assumed that PV trips in the morning start from a traveler's home and trips in the afternoon end there as well. Additionally, it is assumed that parking at home is free and only parking fees for a constant duration in the intermediary location have to be paid. Parking fees depend on the zone where a PV is parked, whereas for the sake of simplicity it is assumed that parking fees are only introduced in the inner city zones z_I, while there are none in the outer part z_O. For the model of this case study, parking fees are equally split across both trips; therefore, the destination of morning trips and the origin of afternoon trips determine the amount:

$$C_{od}^{park,PV}(\tau_i) = \begin{cases} \alpha^P & \tau_i < \text{12:00 } \& \ d \in z_I \\ 0 & \tau_i < \text{12:00 } \& \ d \in z_O \\ \alpha^P & \tau_i \geq \text{12:00 } \& \ o \in z_I \\ 0 & \tau_i \geq \text{12:00 } \& \ o \in z_O \end{cases} \tag{4.22}$$

τ_i reflects the time a traveler wants to start her trip and α^P the regulator decision variable for parking regulations controlling the parking fee rate.

Road Toll Model: Similar to the work of [Bracher and Bogenberger, 2017], a dynamic toll controlled by an NFD is applied. A distance-based toll is used for this case study, where only the distance within the toll area in the inner city $d_{od}|_{z_I}$ is relevant. The coefficient is 0 up until a certain vehicle density threshold k_0. For densities above this threshold, the toll coefficient increases linearly with the vehicle density in the inner city zone and is scaled by the regulator decision variable α^{RT}:

$$C_{od}^{toll}(t) = \max\left(\frac{k_{z_I}^t - k_0}{k_0}, 0\right) \cdot \alpha^{RT} \cdot d_{od}|_{z_I} \tag{4.23}$$

where $k_{z_I}^t$ refers to the density in the inner city zone.

PT Frequency Model: The frequency of PT lines impacts the transportation system in multiple ways. In this model, it affects the total capacity of the PT system and thereby the crowding η_t^{PT} of PT vehicles. Additionally, a higher frequency improves traveler's acceptance of transfers c^T because the average transfer and wait times decrease. For simplicity, the model includes a linear relationship between frequency and the coefficient in the traveler logit model. The effect of scaling the PT frequency by a factor of α^{PT} is modeled as follows:

$$\eta_t^{PT} = \frac{\eta_{t,0}^{PT}}{\alpha^{PT}} \tag{4.24}$$

$$c^T = \frac{c_0^T}{\alpha^{PT}} \tag{4.25}$$

where the "0" lower-case index refers to the value from the original data.

Of course, higher PT frequencies require operating more PT vehicles and generate higher costs and emissions. With the assumed linear relation between number of trips n_l^t on line l per time period and costs in equations (4.20) and $(\mathbf{??})$, following cost scaling can be derived:

$$C_T^{PT} = \alpha^{PT} \cdot C_{T,0}^{PT} \tag{4.26}$$

$$E_T^{PT} = \alpha^{PT} \cdot E_{T,0}^{PT} \tag{4.27}$$

AMoD Fleet Size: The regulator can set an upper bound α^F on the number of AMoD vehicles is allowed to operate. The AMoD operator then has the choice to select the fleet size β^F between 0 and α^F.

AMoD Fare Model: As in the chapter 3, the AMoD system is assumed to use a minimum base fare and a distance-based fare as described by equation (3.1). As mentioned, this service design is chosen to avoid the substitution of short active mode trips by AMoD. Moreover, the utilization-based pricing scheme from equation (3.54) is applied. Together, the fare f_r for a request r is determined by

$$f_r^{AMOD}(t) = \max\left(f^B, \beta^{PD} \cdot d_r \cdot \begin{cases} 1 & \eta_t^{AMOD} < 0.75 \\ 1 + \beta^{PU} \cdot (\eta_t^{AMOD} - 0.75) & \eta_t^{AMOD} \geq 0.75 \end{cases}\right) \tag{4.28}$$

where β^{PD} and β^{PU} are the AMoD operator variables for the distance fare rate and the maximum surge price factor, respectively. It should be noted that the dynamic pricing factor β^{PU} only takes effect if the fleet is utilized to a high degree, whereas the distance-based pricing factor β^{PD} affects almost every single offer and thereby the mode-choice decisions and the transportation system as a whole.

4.3.6 Social Welfare and Profit Model

The objectives of regulator and AMoD service planner are social welfare and profit, respectively. The profit can be defined straightforwardly as the difference between revenues from fares of served travelers and operating costs:

$$P = \sum_{r \in R^s} f_r^{AMOD} - \sum_{v=1}^{\beta^F} \left(c_v^F + c_v^D \cdot d_v\right) - \sum_t C_t^{toll,AMOD} \tag{4.29}$$

where R^s is the set of served requests, f_r^{AMOD} the fare paid by request r, c_v^F the fixed costs of a vehicle (e.g., for leasing and insurance), c_v^D distance-dependent operating costs, d_v the total driven distance of vehicle v and C_t^{toll} is the total toll derived from the distance driven in the toll area of all fleet vehicles ($d_t^{AMOD}\big|_{z_I}$) during a time step with high traffic demand:

$$C_t^{toll,AMOD} = \max\left(\frac{k_{z_I}^t - k_0}{k_0}, 0\right) \cdot \alpha^{RT} \cdot d_t^{AMOD}\big|_{z_I} \tag{4.30}$$

The definition of social welfare is more complex and not unique. Every city could weigh different objectives differently or have additional objectives. The objective function in this case study includes i) the sum of travelers' utilities of their chosen mode $m*$, ii) revenues and costs for PT, iii) revenues from parking and tolls and iv) emissions of the transportation system.

$$\begin{aligned} W = &\sum_r U_r^{m*} + \sum_{r:m_r=PT} f_r^{PT} - C_T^{PT} \\ &+ \sum_{r:m_r=PV} C_{od}^{park,PV} + \sum_{r:m_r=PV} C_{od}^{toll,PV} + \sum_t C_t^{toll,AMOD} \\ &- c^{CO2}\left(e^{PV} \sum_{r:m_r=PV} d_i^{od} + e^{AMOD} \sum_{v=1}^{\beta^F} d_v + E_T^{PT}\right) \end{aligned} \tag{4.31}$$

4.4 Case Study: Setup

4.4.1 Case Study Description

As in Section 3.6.2, this case study is based in Munich, Germany. The AMoD operating area / study area, the street network, and the set of possible AMoD access points for pick-up and drop-off are exactly the same. However, the street network is divided into two zones, one for the city center z_I and one for the outer part of the city z_O, which are separated by the inner highway belt B2R called "Mittlerer Ring" as illustrated in Figure 4.3a. It should be noted that the street network exceeds the study area as trips might be faster along the outer motorway called A99 as through the city. The A99 with a speed limit of predominantly 120 km/h, the B2R with speed limit of predominantly 60 km/h and motorways leading towards the city center with speed limits between 60 and 120 km/h are highlighted in the street network as major streets (see Fig. 4.3a). Most other streets have a speed limit of 50 km/h (general) or 30 km/h (residential areas).

A multi-modal model is required for this framework, especially since Munich has a widely used PT system. The PT mode share in Munich is approximately 40 % within the study area (operating area of AMoD). Hence, the PT network is created based on open-access GTFS data[5]. The included PT lines and stations are shown in Figure 4.3b. The high-capacity trains and corresponding stations in the city center are often overly crowded resulting in stations which act as bottlenecks in the system. Travelers can access or leave the PT network at the PT stops marked in green. To connect the street and PT networks, each PT stop is connected

[5] www.gtfs.de

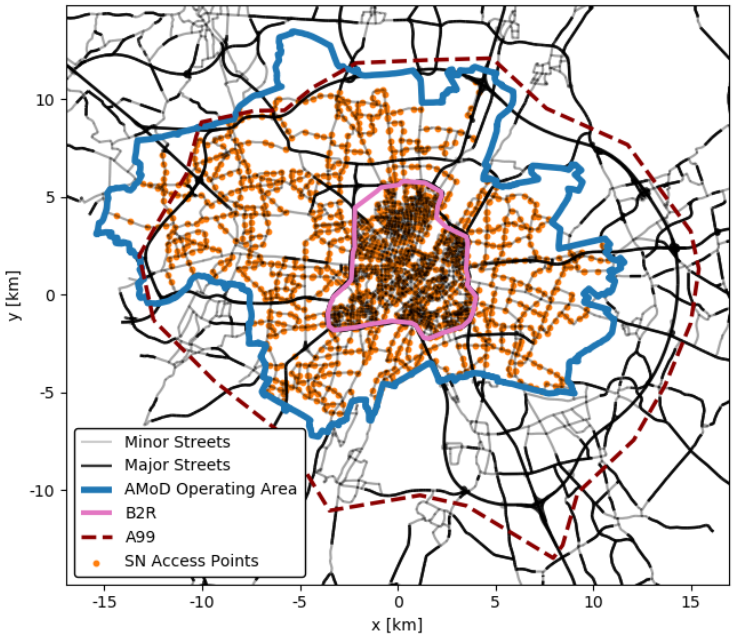
to the nearest street node. Additionally, each street network access point, on which demand is defined, connects to the nearest PT network stop. This simplification is valid as the PT stop density is very high and is used to speed up the pre-processing of the public transport network.

NFDs and functional forms $q_z(k_z)$ are estimated for the street networks within z_I and z_O based on data from the microscopic traffic simulation model. Moreover, the parameters for the travel time scaling behavior in equation (4.16) are estimated. The microscopic traffic simulation model from [DANDL, BRACHER, et al., 2017] also includes PV trips passing, entering, or leaving the A99 area. These trips are treated as background traffic as they are not represented in the traveler list. To calculate background traffic density the model creates two additional traveler sets. The first one includes all PV trips available in the data, while the second one only includes PV trips with origins and destinations within the AMoD operating area. The simulation is run with both traveler sets while enforcing the travelers to always choose PV. The difference between the measured hourly averaged network densities for each NFD-area is considered the background traffic density.

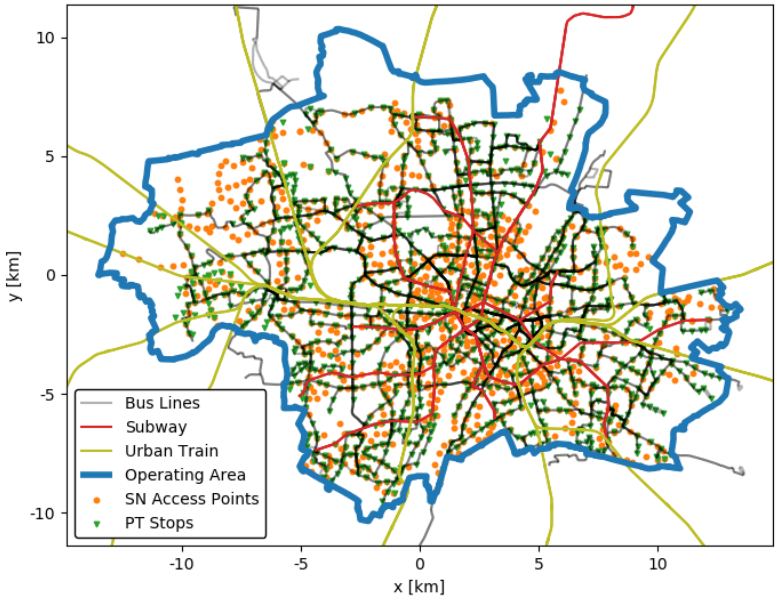
The time-dependent PT capacity is calculated using the GTFS data by evaluating equation $(??)$. The PT background travelers (leaving or entering the study area) are determined the same way as previously described for PVs. Two traveler sets are created based on the original PT demand data from [MAGET et al., 2019]: one includes all PT trips in the data and the other one only includes travelers with both origin and destination within the AMoD operating area. A simulation for each traveler set is run while enforcing the travelers to chose PT. The hourly average difference in PT travelers is used in the model as background value for calculating equation (4.17). Test simulations showed that the crowding factor in the PT system reached a level of approximately 0.2. In the real-world PT system, central stations are highly crowded while many buses, especially in the outer regions, are rarely crowded. In order to model this inhomogeneity in Munich's PT system, where a large share of travelers experience crowding, the total capacity computed according to equation (4.18) is reduced by a factor of 3.

Like in the last chapter, the OD demand data is derived from both data sources ([DANDL, BRACHER, et al., 2017] and [MAGET et al., 2019]), but this time 100 % of demand is represented by agents that want to travel within the study area. The resulting time-dependent trip counts are shown in Fig. 4.4. The travelers are generated by Poisson processes with rates corresponding to the respective OD-matrix entries. The generation process is conducted multiple times to create request data sets with different random seeds. Street network access points within a particular origin and destination area are chosen randomly as traveler origins and destinations, respectively. These agents use the logit model described in Section 4.3.2 for combined mode- and route-choice decisions. By treating travelers from outside the A99-area as background traffic, these travelers are effectively exempted from mode choice decisions. As these travelers do not have direct access to the AMoD service, it is assumed that their decisions are not likely to change thereby reducing the computational effort of the model.

Input parameters for the various sub-models of the transportation model are summarized in the appendix C. Thereby, the value for c^{PV}, the intercept term for private vehicles in the mode choice model, is a result of a calibration process, in which this term was varied in simulations with 0 AMoD vehicles to end up with the ratio of 56.5 % PV and 43.4 % PT (which is known



(a) Street Network



(b) PT Network

Figure 4.3: Street network (a) and PT network (b) of the city of Munich used in this case study. Blue shows the AMoD operating area and orange points correspond to street network (SN) access points, where travelers enter and leave the simulation. Road tolls and parking fee regulations are applied in the "iB2R" area with the "B2R" as outer boundary (a). The area "iB2R" and the remaining study area denoted by "LHM-iB2R" also resemble the two network clusters for calculating the NFDs. In (b) PT stops area shown in green. To connect the PT network and the street network, each PT stop is connected to the nearest street network node.

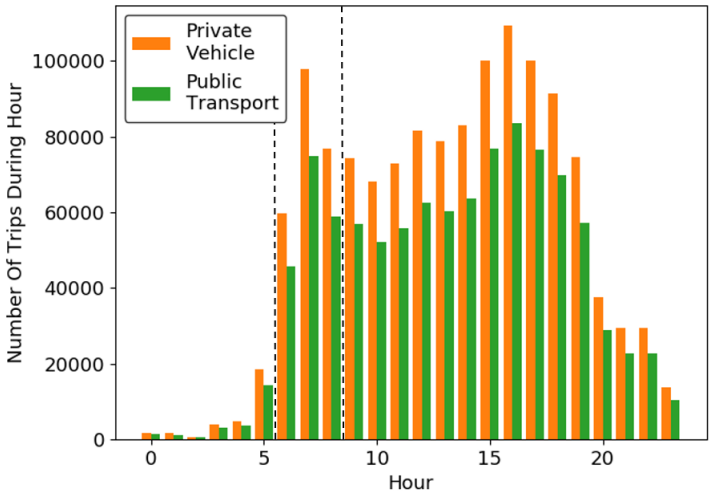


Figure 4.4: Hourly trips within AMoD operating area from PT and PV trip data. Overall the data covers 2.3 million trips with a mode share of 56.6% PV and 43.4% PT. Because of computational complexity, the analysis only simulates the morning peak from 6am to 9am – indicated by the dashed lines. The morning peak contains 18% of all trips.

from the two data sources).

The AMoD service design is chosen as in the operational case study, with constraints for maximal customer waiting time $\tau_r^w = 30$ s and relative detour time $\tau_r^\delta = 40$ % and a vehicle capacity $\rho = 4$ passengers. Additionally, the boarding time is assumed $t^b = 30$ s just as before.

4.4.2 Scenario Setup

In order to improve the livability in the city, the administration and regulator might apply various policies to regulate traffic in a future with AMoD fleets, especially in the inner districts. The case study incorporates the policies described in section 4.3. Table 4.1 shows upper and lower bounds for the corresponding regulator and AMoD service planner decision variables based on discussions with experts from the city administration. These boundaries should reflect socially acceptable limits as, for example, an unbound optimization of the toll parameter might result in a situation, where only the richest people can afford to travel by private vehicle. The scenario "No Regulation" denotes the set of regulator variables $\alpha^P = 2.50$ €, $\alpha^{RT} = 0$ €, $\alpha^{PT} = 1$, and $\alpha^F = 50{,}000$ vehicles. Test simulations showed that this value for α^F is sufficiently high to not constrain the profit-optimal operator fleet sizes.

For application in the Gaussian Process and framework, the variable ranges are linearly transformed to $[0, 1]$ in order to have comparable scales. Mathematically, these limits act as boundaries in the optimization process of the acquisition function in equation (4.9).

With a chosen set of decision variables, the transportation model is simulated three times in the morning-peak (6am-9am) with different random seeds. The resulting mean and standard deviation of W and P are used for μ and σ in the upper level BO framework.

Decision Variable	Player	Range	Unit	Short Description
α^P	R	2.50 - 5.00	€	Parking fees for PVs ending/starting within toll area (morning/afternoon)
α^{RT}	R	0 - 1.00	€/km	Dynamic road toll per driven km for AMoD and PVs within toll area
α^{PT}	R	0.25 - 2.0		Scale factor to decrease/increase PT frequencies
α^F	R	1 - 50,000	vehicle	Number of AMoD vehicle licenses
β^F	FO	0 - α^F	vehicle	Number of AMoD vehicles
β^{PD}	FO	0.25 - 2.00	€/km	Distance-based AMoD fare
β^{PU}	FO	1.0 - 10.0		Scale factor to increase fares in times of high vehicle utilization

Table 4.1: Decision variables of regulator (R) and AMoD fleet operator (FO) in this case study.

Second Social Welfare Definition (Pro-PT Scenario): A second definition of social welfare will be used in the case study to show impacts of other weights of this multi-objective problem. This scenario will be denoted by *Pro-PT* scenario. Following adaptions are assumed:

$$c^{CO2} \rightarrow 25 \cdot c^{CO2} \tag{4.32}$$

$$C^{park,\text{receive}} \rightarrow C^{park}/4 \tag{4.33}$$

$$C^{toll,\text{receive}} \rightarrow C^{toll}/4 \tag{4.34}$$

$$C^{PT} \rightarrow C^{PT}/10 \tag{4.35}$$

$$E^{PT} \rightarrow E^{PT}/10 \tag{4.36}$$

$$e^{AMOD} \rightarrow 3 \cdot e^{AMOD} \tag{4.37}$$

The modified social welfare definition puts a much higher weight on the emissions of the transportation system. Moreover, the standard parameter definition assumes that all parking revenues are paid to the city, thereby ignoring private parking garages. The valuation of parking revenue can also be lower to incorporate the space that has to be reserved for parking instead of put to other productive uses. Similarly, the default definition ignored costs related to the operation of a toll system, a valuation that is given for the required space and infrastructure costs for roads. Therefore, the full parking costs (PV users) and toll costs (PV users and AMoD operator) have to be paid, but only one forth is added as revenue for the city administration. The costs for PT could be lowered with automation and the valuation of its costs could also be lower as PT has the obligation to serve travelers, which naturally will lead to higher costs, but could also be valued. A full electrification, right-sizing, and adaption of the schedule could bring significant reductions in PT emissions. Finally, the AMoD operator might use internal combustion engine vehicles and have an increased CO_2 footprint.

A first impression to sensitivities of different social welfare parameters can be checked very quickly as these coefficients do not affect the input of simulations. Therefore, all simulations

are still valid and the result database from simulations with other social welfare coefficients can be utilized as long as the database contains values for the respective single social welfare components.

4.5 Case Study: Results

4.5.1 Evaluation After Initial Simulations

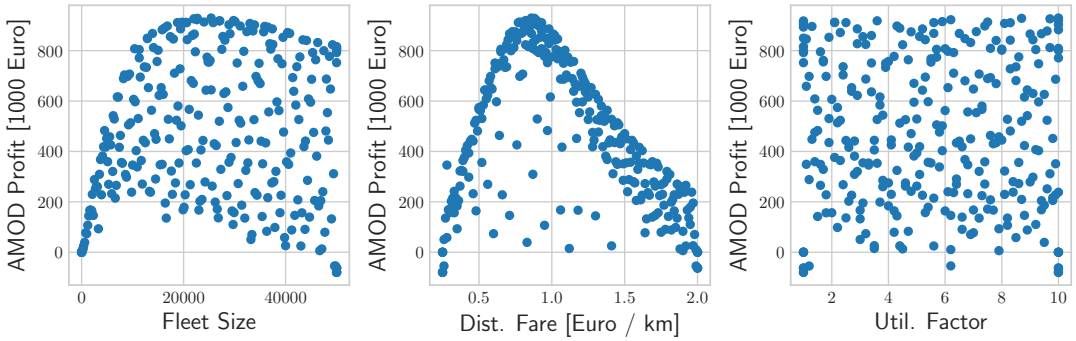


Figure 4.5: Operator profit over operator variables after the initial set of operator simulations for the case without regulation.

An initial set of 272 operator settings for the no-regulation case revealed that the two operator variables fleet size β^F and distance-based fare β^{PD} are more critical than the utilization-based pricing factor β^{PU}.

Figure 4.5 illustrates a sharp increase in operator profit for fleet sizes below 10,000 vehicles. In this region, each addition of an AMoD vehicle generates much more *variable profit*, i.e., revenue minus variable distance-dependent costs, than its fixed vehicle costs. The marginal benefit of adding vehicles decreases up to fleet sizes of 20,000 to 25,000 vehicles before the additional fixed costs of adding vehicles finally become larger than the extra variable profit. Interestingly, the curve does not show a steep decline of profit for larger fleet sizes. This means that vehicles do not have to be utilized the whole time to be profitable. It can be expected that full-day simulations would create a steeper curve since fewer vehicles would be able to serve the demand in off-peak hours.

The distance fare shows a clear maximum at about 0.85 € per km (of direct customer route distance), which at least is in the range of the assumed value of 1.00 € per km in the Chapter 3. In the used mode choice model, fares below that hardly affect the attractiveness of AMoD compared to PT and PV. Hence, higher fares generate larger AMoD revenues. Larger fares reduce the number of AMoD users more than additional revenue per user can bring the operator. Likely, reduced pooling efficiency for a lower number of customers also plays a role. It should be noted again that this case study looks at long-term impacts and computes the PV mode with full costs rather than operating costs. It is probable that AMoD operators have to offer lower fares initially in order to convince travelers to give up their PVs.

As illustrated in Figure 4.6, no more than 12,000 of the more than 25,000 AMoD vehicles are in use at any given time. Only a few of these drive without any passengers, most with

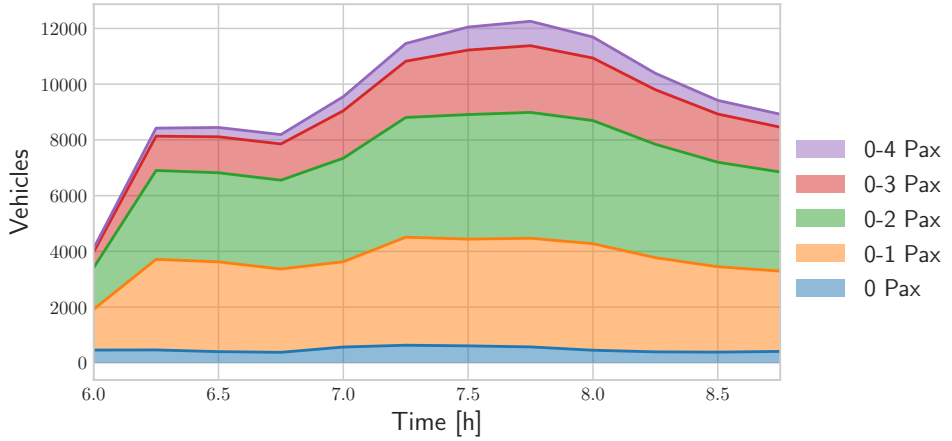


Figure 4.6: AMoD vehicle occupancy in the best operator scenario with 25,400 vehicles and a distance fare of 0.86 € per km.

either one or two passengers, a considerable share with three passengers, and a small share with the maximum of four passengers. This results in a km-averaged occupancy of 1.9 passengers/vehicle. The average travel time and fare of customers are approximately 11 minutes and approximately 3.3 €, respectively.

A simple estimation for the lower bound of profitable vehicle utilization can be conducted with these quantities. With 1.9 passengers on board, a vehicle produces revenues and variable costs of 3.8 and 0.25 € per km, respectively. Therefore, 93 % of revenues are variable profit. Three trips with this variable profit are sufficient to compensate for the fixed vehicle costs. Hence, it can be concluded that the fixed vehicle costs are compensated with approximately half an hour of activity.

As shown in the previous chapter, real-world operators are likely to spend more computational resources and utilize demand estimations to improve the fleet control by more advanced customer-vehicle assignments and repositioning. This will reduce the fleet size needed to serve the same amount of demand thereby reducing fixed vehicle costs. Empty travel and variable costs might increase, but to a lesser degree. For the purpose of this case study, the over-supply of vehicles is acceptable since for the choice of travelers it is critical that vehicles are available; a customer does not consider whether the AMoD vehicle was idle or had to be repositioned beforehand.

The utilization-based pricing factor β^{PU} only makes a difference for fleets in under-supply or near-to-under-supply conditions. Unless the regulator sets a low boundary on fleet size, this will not be the case in the subsequent scenarios.

After the simulations of the initial set of regulator settings, correlations were used as a sanity check for the sensitivities of (regulator) variables in the transportation model. Figure 4.7 displays the Pearson correlation coefficients between several KPIs and specific variables as well as between KPI pairs.

As expected, the modal split of the AMoD service correlates positively with the fleet size and negatively with the base fare. The total traveler utility increases with higher AMoD availability and lower fares. A better AMoD offer also decreases crowding in the PT system and the PV

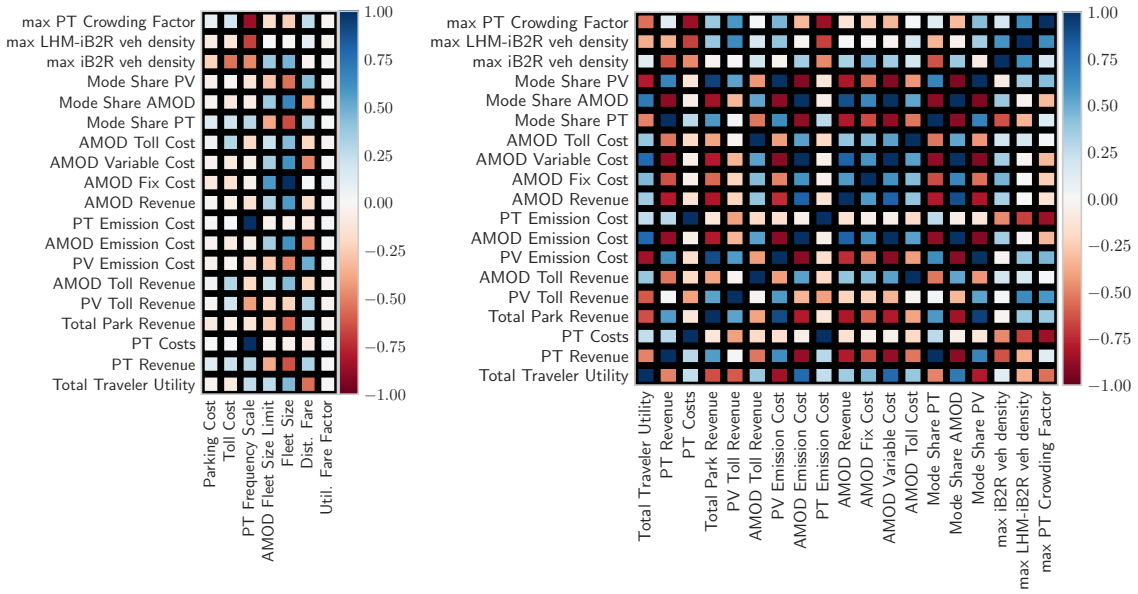


Figure 4.7: Pearson-Correlations of various KPIs with the decision variables (left) and other KPIs (right).

mode share and thereby parking revenues. Obviously, vehicle fixed costs completely correlate with fleet size. Since a fleet size limitation affects the fleet size for low values, its correlations follow the fleet size, but are weaker.

An increased PT frequency effectively decreases crowding, and therefore increases the total utility of all travelers and the modal split of PT. As expected, an increase in PT mode share also reduces the density in both clusters inside (iB2R) and outside (LHM - iB2R) of the inner highway belt (B2R) of the street network.

Parking costs and toll costs generally seem to have smaller impacts than changing the PT frequency and the available AMoD offer (fleet size and fare). Nevertheless, they both serve their purpose increasing the modal share of PT and decreasing the density within the inner city street network.

Most KPI correlations are clear. For instance, low PT emissions are the direct consequence of low PT frequency. Similarly, AMoD fixed costs are a direct consequence of the fleet size variable and therefore, show the same correlations.

Interesting correlations include a negative correlation of PT revenue and total traveler utility. Push measures, which intend to increase the PT mode share to make PT relatively more attractive, increase the general travel costs and therefore decrease the total travel utility. Moreover, high PT mode share is related to scenarios with low AMoD revenue (either due to low fleet size or high fares), which also decrease the total travel utility. All KPIs indicating a high mode share of AMoD (including a low PV mode share) seem to have a positive impact on travel utility.

While correlations are meaningful to derive first insights and perform qualitative model checks, they cannot give quantitative answers to questions, such as how high a regulator should set the toll. Moreover, the correlations do not account for the structure of the tri-

level optimization problem; rather the correlations consider every simulation equally no matter whether the AMoD service planner variables are optimal or not.

4.5.2 Convergence to the Optimal Solution

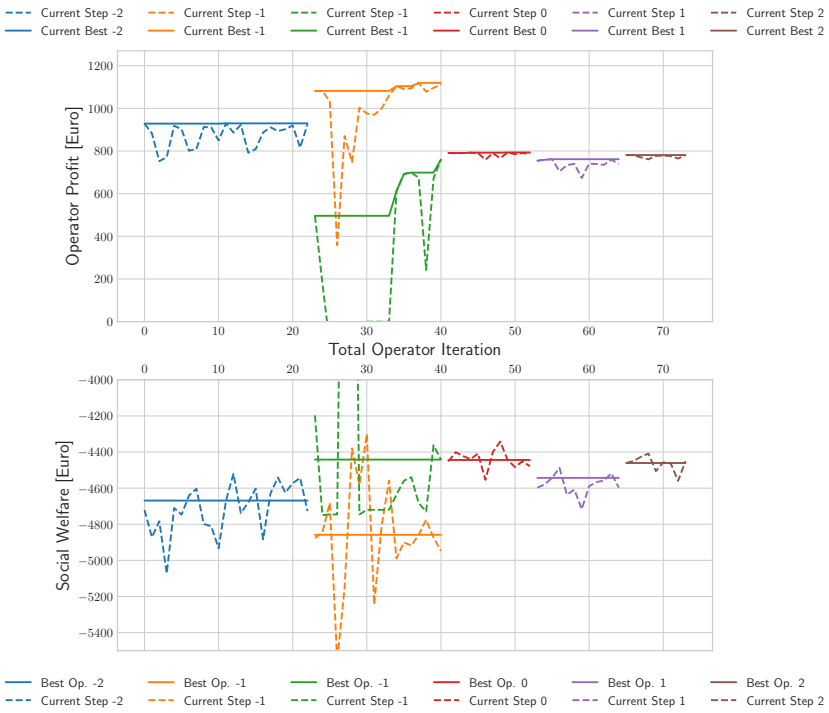


Figure 4.8: Convergence behavior of two-level BO framework. Different colors represent different regulator settings, for which an operator variable optimization is performed. The "-2" iteration represents the initial no regulation setting and the "-1" iteration contains the initial set of regulator settings. The two colors for this iteration show the regulator setting with the best operator profit (orange) and best social welfare (green) result.

The solution approach contains multiple convergence processes as the lower level operator optimization is supposed to converge for each regulator step. Figure 4.8 illustrates the convergence behavior of the two-level framework plotted on a single axis representing the continuously counted operator iteration.

The iteration labeled "-2" denotes the no-regulation case. Its first operator iterations contains 272 different operator settings. After that, each operator step created one new operator setting according to the acquisition function optimization in equation (4.8). A minor improvement over the best initial setting was achieved.

After that, an initial 272 regulator settings were set up, where the respective operator variables were derived via BO. If the fleet size limit was above the optimal fleet size, the best operator variables could be used, for smaller fleet sizes, other operator variable settings were derived. As shown in Figure 4.8, the first operator iterations of the initial regulator

iteration ("-1") often produced really low profit values because the algorithm explored low AMoD fleet sizes or too high or low distance-fare values. The search hardly exploited the prior knowledge from the no-regulation case with κ according to equation (4.10), which motivated the limitation of κ to 1 after 34 operator iterations. With this change, the exploitation of prior simulation results helped to quickly derive near-optimal operator settings.

The large and well distributed set of initial regulator settings in the "-1" iteration ensures that a good solution in the regulator variable space is available.

4.5.3 Comparison of Scenario Results

There are different mindsets when it comes to regulating a transportation system. In the end, the definition of the social welfare function will determine what the optimal scenario will be. In the following, the results from several scenarios are compared:

- *No AMOD*: status quo without an AMoD system
- *No Regulation*: transportation system with AMoD system without changes in regulation; operator variables defined by result of one-level BO
- *Default Scenario*: social welfare definition according to equation (4.31); operator and regulator variables defined by result of BO process
- *Pro-PT Social Welfare Definition*: where the weight of social welfare components are changed according to equations (4.32)-(4.37); operator and regulator variables defined by result of BO process

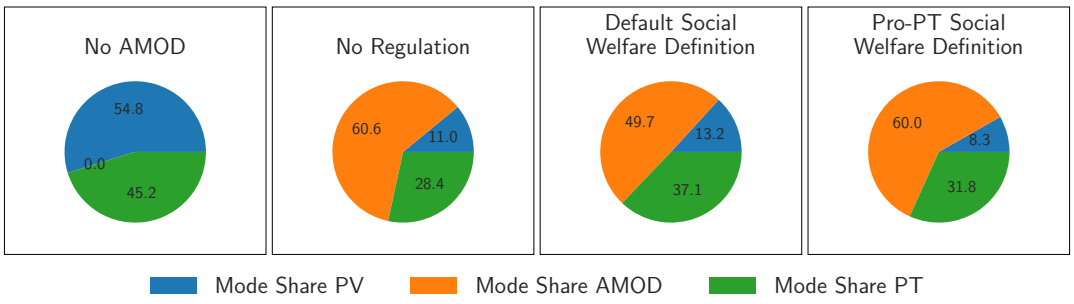


Figure 4.9: Mode split of motorized trips in the described scenarios.

As mentioned in the introduction, AMoD has the potential to disrupt urban mobility. Figure 4.9 illustrates that AMoD will take shares from both PV and PT. In all AMoD scenarios, AMoD was chosen by approximately 50-60% of travelers, whereas a larger share of AMoD travelers were prior PV than PT users. Regulations can impact whether travelers switch from PT to AMoD or PV to AMoD. Interestingly, the two SW definitions introduced measures with quite different outcomes: the standard definition limited the AMoD share at around 50 % thereby increasing both PV and PT share; the Pro-PT definition left the AMoD share unchanged, but further limited the PV share from 11 % in the unregulated scenario to 8 %.

		No AMOD	No Regulation	Default Social Welfare Definition	Pro-PT Social Welfare Definition
Regulator Variable	Parking Cost [Euro / h]	2.5	2.5	2.5	4.84
	Toll Cost [Euro / km]	0	0	1	0.06
	PT Frequency Scale	1	1	0.25	1.23
	AMoD Fleet Size Limit	50000	50000	50000	46900
Operator Variable	Fleet Size	0	25400	14600	22900
	Dist. Fare [Euro / km]	0.25	0.86	1.01	0.84
	Util. Fare Factor	1	10	4.3	10
Social Welfare KPIs	Social Welfare	-5400	-4669	-4442	-4666 / -4727
	Total Traveler Utility	-5331	-4402	-4751	-4359
	ΔTotal Traveler Utility	0	929	580	972
	PT Revenue	178	113	149	127
	PT Costs	334	334	76	405 / 40
	Total Parking Revenue	173	18	15	7 / 2
	AMoD Toll Revenue	0	0	211	31 / 8
	PV Toll Revenue	0	0	33	6 / 1
	PV Emission Cost	32	8	9	6 / 142
	AMoD Emission Cost	0	2	2	2 / 159
	PT Emission Cost	54	54	12	66 / 165
AMOD Profit KPIs	AMoD Profit	0	930	760	899
	AMoD Revenue	0	1365	1275	1348
	AMoD Fixed Costs	0	193	111	174
	AMoD Variable Costs	0	242	193	243
	AMoD Toll Costs	0	0	211	31
Traveler KPIs	Avg. Travel Utility	-12.86	-10.63	-11.47	-10.53
	Avg. AMoD Fare	0	3.3	3.08	3.25

Table 4.2: Regulator and operator variables (section 1 and 2), objective function components (section 3 and 4), and per traveler average values for the described scenarios. The left values in the Pro-PT scenario reflect the original coefficients while the right use the definitions from equations (4.32)-(4.37). The units of the per traveler average values are € and thousands of € for all social welfare and profit component rows are thousands of €.

Table 4.2 summarizes the operator and regulator variables and components for all scenarios. The first observation is that adding AMoD as an additional mode increases the total travel utility considerably, which is also reflected in the large AMoD mode shares. These require large fleet sizes of more than $10,000$ vehicles. The fleet size is not constrained by the best regulatory solution in either regulation scenario, i.e., $\alpha^F > \beta^F$. The fixed costs for the AMoD operator are lower than the variable costs and only a fraction of the revenue resulting in a high profitability of an AMoD service. The AMoD service performance is similar in all three scenarios (with AMoD system). The average waiting time is 130 seconds in the No Regulation and the Pro-PT SW scenario and 140 seconds in the Default SW scenario, which can be attributed to the smaller fleet size. For the same reason, the share of travelers with an AMoD

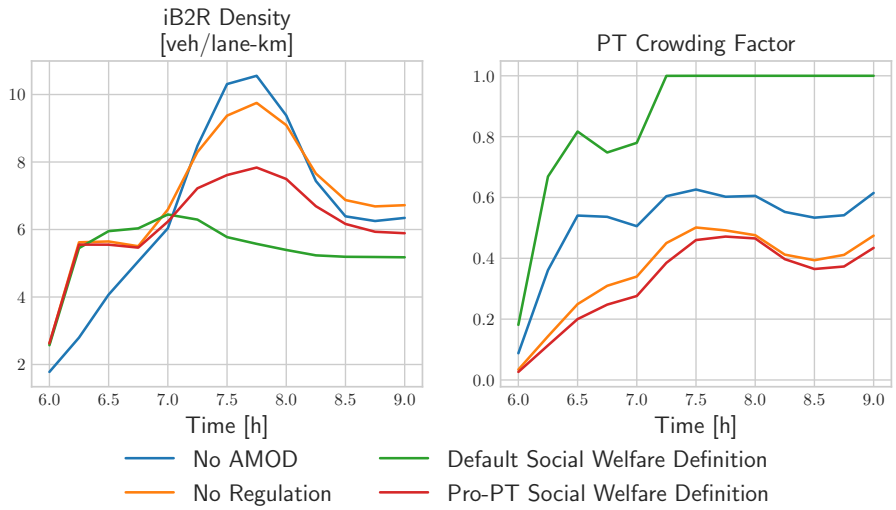


Figure 4.10: Evolution of street and PT network utilization over time in the described scenarios.

offer, which was made if there is a feasible vehicle tour keeping both waiting time and detour time constraints, is slightly lower for this scenario. The average relative detour is between 17 and 18 % in all cases, which in absolute values is in the same range as the waiting time. Since repositioning was excluded in this part of the thesis, the share of empty VKT is less than 4 % in all AMoD scenarios, which is a very low value. On the downside, large vehicle numbers are necessary to cover the area thereby resulting in rather low fleet utilization values between 40 and 55 % in the three AMoD scenarios. Therefore, the utilization surge pricing factor, which would increase fares from 75 % utilization is not relevant. The km-weighed average occupancy is 1.9 in all scenarios; this values represents a considerable improvement over the average occupancy of 1.0 for PVs[6]. With a large share of AMoD users originating from the PV sector, the densities in the street network are reduced even without regulation (see Figure 4.10). Moreover, parking revenues are drastically decreased compared to the No AMoD scenario, which also can be viewed as positive since parking space is freed for other use.

The standard definition of social welfare employs an emission cost coefficient from the German Plan for Federal Traffic Routes. The contributions with this definition are an order of magnitude too small to effectively steer the regulatory policy. Assuming a fully electric fleet and considering only the emissions for the production of the current (instead of life-cycle considerations for simplicity), the introduction of an AMoD system is clearly beneficial over the No AMoD scenario representing the status quo. The combined emissions of PV and AMoD are reduced by approximately 69 %. On the contrary, the PT system in Munich in its current state is neither economical nor ecological. Besides the city center, there are many low occupancy trips, which produce both high operating and emission costs. Especially the operating costs are a substantial contribution to the social welfare. Hence, the regulator in the default social welfare definition reduces the PT frequency to the minimal allowed value and increases the toll

[6]As with many transportation models, groups with multiple travelers are not represented.

to its maximal value while leaving the parking costs unchanged. The dynamic toll sets in for densities over 5 veh/lane-km and increases with increasing density and has to be paid by PV and AMoD. Therefore, the high toll in the Default SW scenario is most effective in keeping the density of vehicles low in the inner city. The AMoD operator reacts by decreasing the fleet size and increasing the fare. Interestingly, the fare is only increased by a fraction of the toll, namely by 0.15 € per km. As illustrated in Figure 4.10, the toll is high enough to keep the density in the inner city rather low, but the crowding in the PT system is really high due to the frequency reduction. Both effects are reflected in an decreased value of the traveler utility compared to the unregulated scenario. Since the AMoD price model assumes an increased per-km price also in the outer region, where no toll is applied to PV and additionally trip lengths and AMoD fares are higher, the ratio of PV to AMoD changes towards PV for trips with origin and destination outside of the city center (see last section in Table 4.3).

Origin	Destination	Scenario	Mode Share PV [%]	Mode Share PT [%]	Mode Share AMoD [%]	Avg. PV Park Cost [€]	Avg. PV Toll Cost [€]	Avg. Offered AMOD Fare [€]
iB2R	LHM-iB2R	No AMOD	66.0	34.0	0.0	0.0	0.0	0.0
		No Regulation	16.9	30.7	52.3	0.0	0.0	6.3
		Default SW	14.3	41.8	43.9	0.0	2.6	7.1
		Pro-PT SW	13.8	33.2	53.0	0.0	0.4	6.1
LHM-iB2R	iB2R	No AMOD	39.2	60.8	0.0	2.5	0.0	0.0
		No Regulation	3.2	33.0	63.8	2.5	0.0	6.5
		Default SW	3.2	42.1	54.7	2.5	2.4	7.5
		Pro-PT SW	0.6	36.2	63.1	4.8	0.3	6.3
iB2R	iB2R	No AMOD	34.0	66.0	0.0	2.5	0.0	0.0
		No Regulation	4.6	26.3	69.1	2.5	0.0	3.1
		Default SW	3.0	39.2	57.9	2.5	2.6	3.4
		Pro-PT SW	0.9	33.9	65.2	4.8	0.4	2.9
LHM-iB2R	LHM-iB2R	No AMOD	72.2	27.8	0.0	0.0	0.0	0.0
		No Regulation	16.8	24.3	58.9	0.0	0.0	7.3
		Default SW	26.9	28.4	44.7	0.0	0.7	8.5
		Pro-PT SW	14.9	25.7	59.4	0.0	0.1	7.1

Table 4.3: Analysis of zone-to-zone travel relations in the described scenarios.

The Pro-PT scenario assumes a more efficient operation of the PT system, both ecologically and economically. Moreover, the regulator is assumed to put a much higher weight on emissions. In order to optimize social welfare in this scenario, the regulator asks the the PT operator to increase the frequency and only sets a minor toll while drastically increasing the parking costs. This further improves the space gained by the introduction of an AMoD system. There are hardly PV trips ending in the iB2R region, where PVs have to pay parking fees (see zone-to-zone travel relations in Table 4.3). The AMoD system is operated with similar parameters as in the unregulated case. The densities in the street are lower than in the No AMOD and No Regulation scenario and the PT system shows much lower levels of crowding (Figure 4.10).

4.5.4 Sensitivity of AMoD Operator and Regulator Variables

The AMoD profit surrogate function $P^S(\alpha^y, \beta^s)$ is a 7-dimensional function. Since the AMoD service planner only has control over its own variables, the four regulator dimensions can be viewed as inaccessible. The regulator sets the hyper-planes in which the AMoD service planner can optimize. In the Default SW scenario, the regulator sets a high toll of 1 € per km and decreased the PT frequency. For this regulatory setting, the variable costs are increased by

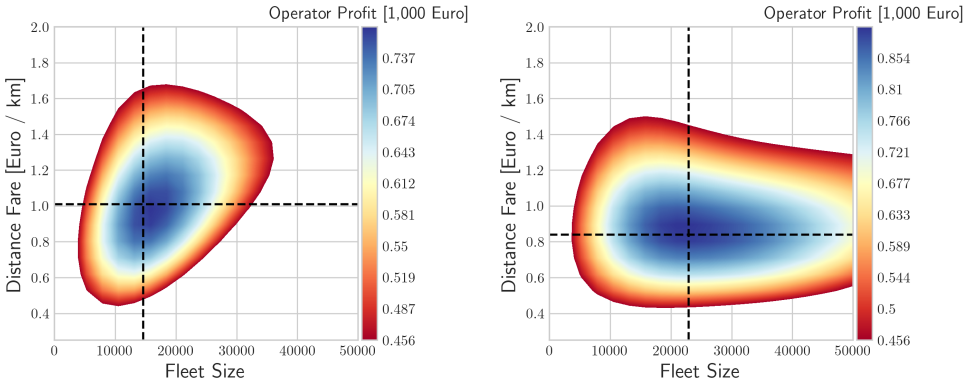


Figure 4.11: 2D cuts of the AMoD profit surrogate function along AMoD operator variables in the hyper-planes defined by the described solutions for the Default Social Welfare Definition (left) and Pro-PT Social Welfare Definition (right). The white area shows the non-profitable area in the solution space and the dashed lines highlight the hyper-planes with the selected AMoD variables $\beta^{s,*}(\alpha^y)$.

a factor of 4. As pooling allows a km-averaged occupancy of nearly 2, the AMoD operator can actually offer the service at fares of approximately 1 € per km and be profitable. However, compared to the Pro-PT SW scenario, in which parking costs instead of toll costs were increased, the area of profitability is drastically reduced (see Figure 4.11).

The 4-dimensional surrogate functions $W^S(\alpha^y)$ are based on less data, since only the simulations with the highest operator profit contribute data points — one for each regulator setting. Figure 4.12 shows some two-dimensional cuts through the hyper-planes defined by the best solution for the two social welfare definition scenarios. The left and right sides clearly paint two very different pictures: while the Default SW function decreases with decreasing PT frequency, the best solution in the Pro-PT scenario is to increase PT frequency. Fleet size limitations below $10{,}000$ vehicles have a very negative impact on social welfare and become minor up above that. As mentioned before, limitations above the profit-maximizing fleet size do not matter.

The real advantage of the surrogate functions is the combined consideration of variables. Assuming a toll of 1 € per km and parking costs of 2.50 €, the worst decision according to this SW definition would actually be to forbid AMoD and keep the PT frequency as it is. As the ecological footprint is weighed highly in the Pro-PT case, both increases in PT frequency and AMoD fleet size (increasing fleet size limit from 0) are beneficial for social welfare as both modes attract travelers from PV users, which have the worst CO_2 footprint.

Parking and toll costs both reduce the travel utility of the PV thereby nudging more people to use PT. A combined increase of both toll and parking costs could be interesting for city administrations. The Pro-PT scenario shows that after a certain point, an increase is not really beneficial anymore because benefits from travelers changing mode do not compensate for the increased total travel costs anymore. At this point, city administrations will also have trouble justifying their measures and run into trouble with social equity.

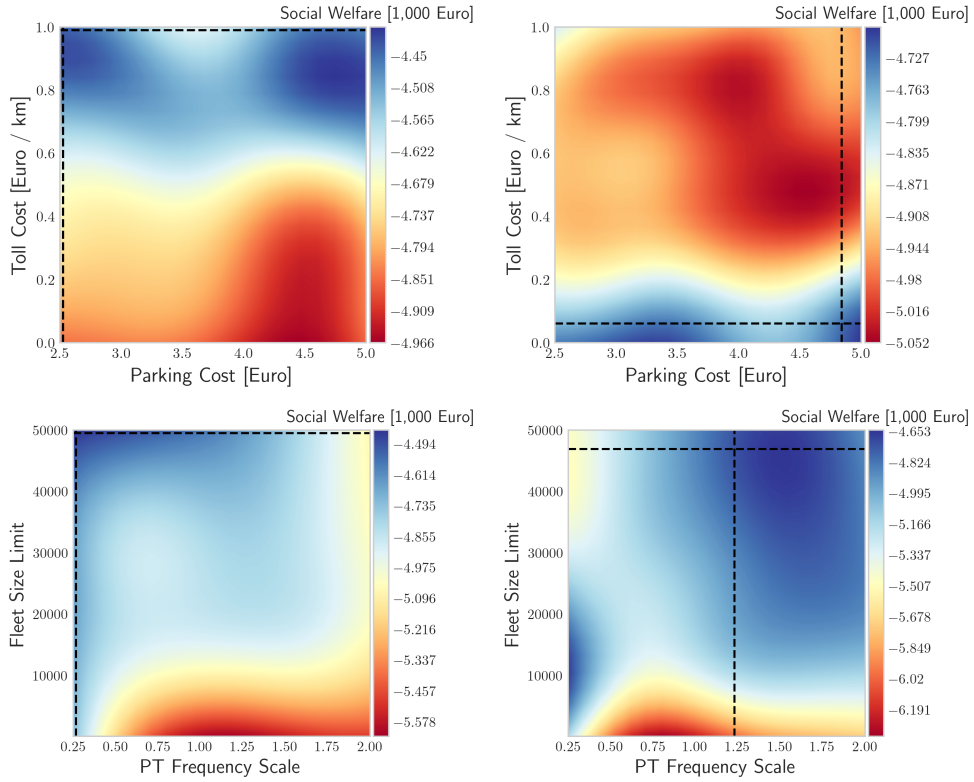


Figure 4.12: 2D cuts of the social welfare surrogate functions along regulator variables in the hyper-planes defined by the described solutions for the Default Social Welfare Definition (left) and Pro-PT Social Welfare Definition (right). The dashed lines highlight the hyper-planes with the selected regulator variables $\alpha^{y,*}$.

4.6 Conclusion

Besides the continuous regulatory decision parameters of the BO framework, the case study contained one implicit regulation: the city administration only allows a pooling-service. This does not mean that all AMoD customers have to be matched; they can also be alone in a vehicle. However, they would be pooled if it is beneficial according to the fleet control objective. The AMoD operator and city administration objectives are in line as both benefit from people sharing a ride: the operator has a better ratio of revenue to vehicle miles and the city benefits from fewer vehicles on the road.

The AMoD ride-pooling system scales very well with increasing demand. AMoD customers only had minor waiting and detour times. On average, travelers had to endure 5 minutes for the sum of both. Considering the time and stress involved in finding a parking space in the city, the level-of-service is comparable to PV. This case study used PV full costs (0.60 € per km) in order to analyze long-term adoption. The AMoD operator adapted the prices according to its competition (0.85 € per km in the scenarios without a toll), which was more than three times its variable cost. Combined with an average occupancy of nearly 2 passengers/vehicle, the profit margin of the AMoD system is very large. It is likely that an AMoD service has to

compete with operating costs in the introductory phase, in which private vehicle ownership is rather high. The optimization of profit would motivate the AMoD operator to offer lower fares than in the studied scenarios. Therefore, the profit margin would be considerably lower. However, in the long term, travelers might be willing to abolish private vehicles if they have a better alternative.

The introduction of an AMoD pooling service brings many social benefits. Travelers gain a new and very attractive mode, which can also help to reduce the number of vehicles on and off the road. Additionally, AMoD is environmentally more sustainable than PV as they are likely to be powered by electric motors and have a higher vehicle occupancy. They even compare favorably against today's PT system in Munich, where large buses drive rather often with only a few passengers.

These benefits caused the policy optimization to not apply a fleet size limitation. This policy recommendation is likely general and transferable to other cities, but has to be analyzed in detail to be certain. Unless other factors not considered in this case study give reason for a limitation, fixed vehicle costs limit the fleet size in the profit optimization. Therefore, policy makers should consider extending the PT portfolio to include an AMoD pooling service as it can be open to the public and does not exclude travelers. In a PT integration process, AMoD should not only be considered as an intermodal feeder service. As mentioned, the large social benefits actually originate from its competitiveness with PV. Test simulations with an intermodal feeder service and the current PT system showed that only a minor share (less than 1 %) of travelers used AMoD to get to the next rail-based PT station. The bus stop network in the study area is very dense and frequencies rather high; travelers rarely chose to wait for an AMoD vehicle and pay extra for AMoD for a first/last mile trip. Due to its high computational burden and minor impact, the intermodal sub-model was not included in the final version of the transportation model. However, an intermodal feeder model should be included in an integrated PT-AMoD planning process, in which AMoD is considered to replace inefficient bus lines and off-peak hour bus services.

It is difficult to derive other general policy recommendations. There will not be one set of regulations that are optimal for every city because measures of social welfare are subject to the preferences of each city and its residents. For example, some cities might weight emissions higher than traveler utilities. There is an effort to derive more general insights by classifying cities and simulating a set of archetype cities [Oke, Aboutaleb, et al., 2019]. However, the generalization still assumes that the social welfare definition and traveler behavior are similar in the same city classes. As shown with the two social welfare definition scenarios, the quantitative values of regulatory measures depend on the exact definition of social welfare. The framework in this chapter requires an explicit weighting of different social welfare components, the effect of which should be checked in order to set meaningful weights. The case study in Munich showed that the coefficient used in the Bundesverkehrswegeplan, a plan for future infrastructure projects by the German Federal Ministry from 2016, is so low that emissions had practically no impact on the policy optimization.

In general, push measures are a useful tool to decrease the modal split of PV. Nevertheless, it should be ensured that the balance between additional traveler costs and social value remains in balance. A concern related to very high push factors is that they can increase social inequity. If the push factors are strong enough, only the wealthy travelers in the region will be able to

use PVs. The framework presented in this chapter is a tool that will stop increasing the value of toll and parking costs when there is no more social gain. Further considerations, e.g., regarding social equity, can help to set additional boundaries on the variables [Bogenberger et al., 2021].

The application of two social welfare scenarios show that AMoD operators will react to policies and adopt their service, which in turn has an effect on travelers' choices. One contribution of this chapter is the development of a framework that allows regulators to consider this AMoD operator reaction in order to derive good policies for their respective social welfare definition. Moreover, the impact of weights in the social welfare definitions can be analyzed using the surrogate functions that are part of the BO procedure.

The framework could integrate more complex mode-choice models to represent travel behavior more accurately. The acceptance of AMoD (especially that it is only offered as a pooling service) and travel behavior are likely the largest uncertainties in the system. More advanced behavioral models require more data to estimate a larger number of coefficients related to travel behavior. Generally, the solution approach to the tri-level problem can integrate more refined models (e.g., fleet control); however, there is direct trade-off between model resolution and computation time in the proposed tri-level model and simulation-optimization solution approach. Regarding AMoD profit, the cost structure of autonomous vehicles represents significant uncertainty. Nevertheless, as long as the AV costs remain below the fares, the prices could be market-driven and the modal split similar to the presented results.

Chapter 5

Concluding Remarks

5.1 Summary

This thesis addresses two main topics, namely the operation and regulation of AMoD systems. These topics will become relevant with the introduction of AVs, which are fully capable and legally allowed to drive on their own. AVs can change the cost structure of MoD services significantly and thereby make AMoD services highly attractive.

Operation of AMoD: Chapter 3 studies how an AMoD service provider can operate its fleet effectively and efficiently. Therefore, a certain AMoD system design is assumed, including the costs, the fares, the offered service quality in terms of maximum waiting and detour times, and the demand such system can attract. The contributions of this chapter are the mathematical formulation of the underlying problem and the development and evaluation of solution approaches to this highly dynamic and stochastic problem. The methodology is based on a separation of time scales, which defines actions based on their respective time horizons and therefore splits the problem into several easier to tackle problems. The evaluation is performed with the help of an agent-based simulation framework and a case study for the city of Munich, Germany. Hereby, the interaction between users and the fleet operator are modeled with several phases: a request by a user is answered by an offer of the AMoD operator, based on which the request determines whether to book the service. If the service is booked, an AV of the operator picks up and drops off the user at her start and end locations, respectively.

Algorithms to create offers and vehicle plans, which determine the routes vehicles follow, are developed in the short-term section for both hailing and pooling. Insertion algorithms are used to derive realistic offers, and global re-optimization of all vehicle plans is applied periodically. Throughout the case study, the positive scaling effects of AMoD services become apparent. With a higher vehicle density, the probability of finding a nearby vehicle that can be matched to a new request is higher, thereby increasing the service rate. There is an additional positive scaling effect with demand for pooling services as the probability of finding matching requests increases. In the case study, the insertion heuristics generally find good solutions, and optimization potential seems somewhat limited. This is in contrast to the results found by M. Hyland and Mahmassani [2018a] but can be explained with the different service designs: the stricter the time constraints, the less optimization potential is available. Here, the maximum waiting time is restricted to 5 minutes, whereas there is no time restriction applied in M. Hyland and Mahmassani [2018a] to model a service guarantee. Hence, dynamic effects even overshadow possible benefits from optimizing the current state in the hailing case.

In the pooling case, there is potential for optimization even with these strict time constraints, but they also come with a much higher computational cost.

The knowledge of future demand is studied both by including reservations and utilization of aggregated forecasts. Having a completely reservation-based AMoD system increases the performance of both hailing and pooling compared to on-demand AMoD systems. However, allowing both online and reservation-based requests even performed worse than a system with only online requests. Even though the reserved requests can be served and optimized well, the space to integrate online requests becomes very constrained, leading to low service rates for the online requests. Repositioning based on demand and supply forecasts helps to position vehicles where they are needed ahead of time. In general, repositioning is a multi-objective problem, in which the imbalance of the system should be equalized with as little vehicle movement as possible. New computationally-efficient multi-step density-based repositioning problem formulations are developed in this thesis. The case study shows significant benefits through the introduction of repositioning. Although compared to that large gain, the performance gains of the new density-based formulations compared to existing approaches seem quite small, it has to be noted that these add up over time, and AMoD providers are likely interested in these improvements in the range of a few percent.

An analysis of the dependency of the fleet size with respect to the cost and fare structure shows that In a profit over fleet size plot, the fixed costs determine the steepness of the decline for larger fleets, whereas variable costs determine the steepness of the rise for smaller fleet sizes. Therefore, increasing fixed costs tends to decrease fleet size more while having less effect on the actual profits when compared with increasing variable costs. AMoD operators can utilize dynamic pricing in times of under-supply to increase profit while keeping similar service rates. The concept shown in this thesis is to set the fare at a level such that only as many users want to book trips as there is supply available.

In general, the assumed cost and fare structure are very favorable for AMoD services. Compared to today's carsharing system, the operating area can be extended beyond the city center to the city boundaries with lower trip densities with increasing profits. The amount of trips an AMoD vehicle can serve is very high; in the case study, more than 115 thousand trips are served by less than $2,000$ vehicles. Therefore, the introduction of an AMoD service offers substantial potential to reduce parking lots. Nevertheless, AMoD vehicles also should be parked when they are not needed for serving requests and in the right place for anticipated demand as otherwise, they might circle around the streets.

Reflecting on the research questions

RQ 1.1 How can an AMoD fleet be operated effectively and efficiently?

RQ 1.2 Which are the most important operational variables?

it can be concluded that this thesis

- provides a detailed agent-based model that describes how the fleet-operational problem can be tackled and an AMoD system can be operated,
- adds contributions to improve operational strategies, especially in the context of repositioning, and

- identifies the long-term strategies for planning the system design parameters as the most relevant operational parameters.

The thesis cannot provide an optimal solution for the dynamic and stochastic fleet-operational problem. Hence, it is possible and likely that better strategies can be found in the future as fleet operators will invest human and computational resources to continue improving these strategies to enhance their performance. The contributions in this thesis can be a step-stone.

The second research question is the identification of the most important operator decisions. Planning the long-term system design parameters are the most relevant for an operator's profit and the share of served requests: whether to offer a hailing or pooling service, setting the service level (time constraints, fares), and choosing a fleet size. Heuristic solution strategies for the more frequent operational tasks (mid- and short-term decisions) offer a good trade-off between solution quality and computational effort. These insights can be used for the integration of an AMoD operator in the larger context of a transportation system.

Regulation of AMoD: While there was a lot of research existing on how to operate a fleet, the question of regulating this fleet in the context of the transportation system was hardly addressed before. Therefore, the research questions

RQ 2.1 What are the effects of an AMoD service on a transportation system?

RQ 2.2 How will an operator react to different regulatory measures?

RQ 2.3 How can city administrations regulate the transportation system considering the reaction of AMoD operators?

require a novel problem formulation and solution approach, both of which are introduced in this thesis.

The contributions of this thesis are

- the introduction of the tri-level AMoD regulation problem,
- the general two-level BO solution approach,
- the development of a transportation model on the third level, which has to find a good trade-off between model resolution and computational complexity, and
- a case study demonstrating the framework in the city of Munich.

The problem considers a separation of scales to build a hierarchy of decision-making. First, a public entity (on state, federal, or municipal level) sets the regulatory frame in which an AMoD provider offers its service. Second, the operator aims to optimize its profit and adapts its service according to the regulatory setting. On the third level, travelers determine their mobility behavior based on the trip characteristics that can be offered by the different modes in the transportation system, which are affected by the decisions made on the upper two levels. Ultimately, the goal of the framework is to provide an optimization framework for the regulating entity, which considers the impact of its decisions on the two lower levels to optimize social welfare.

A two-level BO solution is developed to search the solution spaces of regulator and operator variables. The evaluation of each data point in this solution space involves a simulation of the transportation model, which is a very time-consuming process for an optimization procedure. Compared to a naive two-level optimization with independent optimization loops, the advantage of the BO approach is that surrogate functions can be inferred from all previous simulations. Therefore, approximations of the relevant objective functions defined on the solution space are available for the optimization process and help to reduce the required number of simulations.

Even with this approach, a considerable amount of simulations is necessary. Hence, the strategy for the transportation model is to go into detail where sensitivity to decision variables is required and aggregate where it is possible. Inserting every request into the existing vehicle plans to create an offer is very time-consuming, which limited the simulation period to the morning hours. Nevertheless, this part of the thesis also follows an agent-based modeling approach, because individuals' decision-making remains tractable, and an explicit matching of travelers and vehicles ensures a realistic level of service model for an AMoD system. The spatial resolution of street and PT networks is also very high. Still, their dynamics are included on an aggregate level allowing for a complete pre-processing of the routing sub-problems.

There are a lot of parameters that a regulator and an AMoD operator can vary to influence the transportation system. For the case study, it is assumed that the regulator enforces pooling to avoid the negative externalities caused by hailing. The chosen set of regulatory measures considered for optimization are parking fees for PV, a dynamic toll for PV and AMoD, an AMoD fleet size limit, and changes in PT frequency. The AMoD system utilizes a service design (concerning time constraints) similar to chapter 3 and optimizes the fleet size and fare. As computational efficiency is relevant, the model refrains from re-optimizing user-vehicle assignments, which likely leads to an underestimation of fleet efficiency.

A generalization of the case study results is difficult as the definition of social welfare is not unique. Each city administration might have its own ideas of its goals and how to value certain aspects. Additionally, the transportation networks can have very different characteristics leading to significantly different sensitivities. For instance, a city with a well-functioning PT system will value the existing system highly. In contrast, a city with hardly any PT could see AMoD as a chance to replace highly inefficient parts. Consequently, the shape of the social welfare function in the regulatory solution space and the optimal solution of the regulatory variables depends on the city. Hence, the main contribution of this chapter is the developed methodology that can be applied to different cities.

Some results of the case study are likely transferable. The addition of an AMoD pooling service — even without regulatory measures — increases the travel utility of the population compared to the status quo. A regulatory limitation of the AMoD fleet size does not seem meaningful; the operator will adapt its fleet size to serve the attracted demand anyway. With this new mobility option as a pull measure, regulators can introduce push measures like toll and higher parking fees to steer the mode choice behavior and increase travel utility compared to the status quo.

5.2 Future Research

On the operator side, this thesis showed one method to approach the highly dynamic and stochastic problem. However, it is unlikely that an approach can be found that solves this problem optimally in all cases as long as computational resources are a limiting factor. Solving vehicle routing problems could be a future application of quantum computers. Until then and even beyond, variants of vehicle routing problems are likely to remain an active field of research. Especially different service designs or more realistic frameworks will create new variants of problems, which for example consider intermodal and stochastic networks, mixed fleets, the combination of logistic and people transportation, interactions of an electric AMoD fleet with the power grid, the use of access points, non-instantaneous decision-making by users and operators, late- and no shows, cancellations, and constraints due to information processes to increase user convenience such as locking of vehicle and pick-up time a certain time before the expected pick-up.

Likely, the approach to dividing the action space and addressing the smaller sub-problems will still be a reasonable approach then. Nevertheless, even in the AMoD service design assumed in this thesis, there is still plenty of room for improvement. Especially for tasks with high computational complexity, e.g., the optimization of pooling assignments, the use of access points, or intermodal assignments, the computational complexity gives reason to study heuristics that limit the solution space in meaningful ways. In the context of the framework presented in this thesis, the parameter space of the presented RV heuristics can be analyzed, and additional RV heuristics can be developed. Additionally, the multi-objective aspect in the control function of the assignment processes can be analyzed in more detail. Finally, the expected future vehicle imbalance distribution can also be included as an objective in the assignment process [Syed, Dandl, and Bogenberger, 2021].

The forecast methodology of demand and especially supply can be improved upon. In particular, for pooling, the supply forecast methodology should consider that a vehicle can offer multiple seats, which will likely be available in a specific direction at different times. On the demand side, considering the pooling potential for certain zones or origin-destination pairs represents a possible advancement worth investigating. Even though the forecasts work well for repositioning in the current state, better results can be expected from an improved accuracy of demand and supply imbalances. Furthermore, the presented one-step forecast dynamic pricing would benefit from better predictions.

As mentioned when introducing the original problem in equation (3.4), approximate dynamic programming approaches can be used as alternative solution methods. Machine learning, specifically reinforcement learning, can be used in this context. Even then, researchers will likely split the overall problem into easier sub-problems again to reach practical learning curves. The repositioning problem derived in this thesis can serve as one example where decisions with good future impacts can be learned from the current state and imbalance density forecasts. The density representation of imbalance developed in this thesis might be especially beneficial in combination with convolutional neural networks. Another application for machine-learning in practice could be the creation of realistic offers purely on the current system state and request information without the need for explicit insertions of new stops into vehicle plans. Besides being useful for real-world operators, the feature of estimating realistic offers could

be of great significance for the modeling of AMoD as part of transportation systems as fast AMoD offer estimators would allow extending the simulation period in the same computation time.

The developed tri-level framework to optimize regulations gives rise to several new research ideas. It is noteworthy that the inherent computational complexity of the problem always requires a trade-off between model resolution and model sensitivity towards regulatory variables. An interesting approach would be to use a more macroscopic approach for the transportation model and compare performance and results with the current model.

The current model can be improved by getting more data about traveler behavior, building an activity-based model, and refining the mode choice model. A combined approach about within-day mode-choice and pre-day mode choice could be investigated, in which travelers can also learn during the iterations and some of them can become captive PT, PV, or AMoD users in the transportation model simulation. On the supply side, repositioning can be integrated into the current model to reflect more realistic operator behavior. The difficulty is creating the forecasts of demand and supply, which could be based on the general trip density and prior simulation iterations.

This thesis assumes that a single AMoD provider operates in a city, which leads to very high profit margins of this single provider as it sets the fares to compete against PT and PV. Competition with other AMoD services would likely lead to fares on a lower level closer to the actual AMoD costs with some percent profit margins. The framework can easily be extended to let multiple operators choose their variables in sequence[1]. With multiple AMoD operators, there is typically a price of fragmentation. Different concepts of cooperation are possible to alleviate these losses [Engelhardt, Malcolm, et al., 2022].

Another research direction would be to follow up on Pinto et al. [2019]: why not fully integrate a profitable mobility service, namely AMoD, into public transportation? The current line-based concept works well on connections with loads of demand, especially if it is on rail and does not run into the same congestion as the street traffic. However, many connections follow a downward spiral: little demand is expected; therefore, the service level is low, which again dissuades demand. These connections are also very costly. A combined PT and AMoD planning process should be started from the existing PT infrastructure and lines where and when they work well. The AMoD service service could be offered only for connections not parallel to these lines to avoid cannibalization. One step before prohibiting these trips altogether would be the utilization of pricing; trips in parallel to the reduced number of PT lines should be more expensive. These suggestions could also work as regulations for open-market AMoD services. Important are (i) the PT re-planning process beforehand and (ii) to think of AMoD as more than just a feeder of line-based systems. The inherent advantage is that almost all connections in a large city area become available to a public transportation system with minor detours. This new connectivity is very different from the current PT system, where transfers between lines are often required and lead to large time losses compared to PV. Therefore, there is a significant potential for AMoD services to improve future transportation systems if appropriately integrated.

[1]This has recently been proposed in a publication by Bortolomiol et al. [2021], which has just been published during the writing process of the conclusion and also tackles the problem of regulation with mobility service suppliers.

List of Figures

List of Tables

List of Terms and Abbreviations

AMoD autonomous mobility on demand 1, 4–9, 11–25, 27–29, 31–35, 37–39, 41, 45, 46, 49, 53–55, 68, 70–73, 78–82, 84, 103, 117, 120–126, 128–133, 135–162, 193, 194, 196, 202, 203, 207, 208, 210

AV autonomous vehicle 4–7, 11, 16, 20, 23, 25, 27–29, 35, 73, 111, 121, 122, 155

BO Bayesian optimization 125–128, 141, 146, 147, 152, 154, 157, 158

DaRP dial-a-ride problem 22, 48, 50, 53

IRS immediate response system 83–85, 89, 91, 92, 97, 99

KPI key performance indicator 76, 77, 86, 90, 97, 100–102, 144, 145, 148, 161, 162

MaaS mobility as a service 16, 20, 207

MoD mobility on demand 1–5, 12–16, 18, 20, 25, 27–29, 121, 155

NFD network fundamental diagram 133, 136, 139

OSM OpenStreetMap 17, 18

PKT person kilometers traveled 77, 95, 121

PT public transportation 2, 3, 7, 11, 12, 16, 17, 19, 20, 23, 28, 72, 83, 84, 117, 122, 124, 129–135, 137–145, 147–151, 153, 158, 160, 162, 193, 194, 204, 207, 208

PV private vehicle 83, 84, 124, 129–133, 136, 139, 141–145, 147–154, 158, 160, 194

V2RB vehicle to request bundle 48–52

VKT vehicle kilometers traveled 2, 3, 5–7, 14, 28, 29, 45, 59, 77, 89, 95, 102, 103, 105, 111, 113, 118, 120–122, 130, 136, 149

VRP vehicle routing problems 11, 15, 22, 25

Bibliography

Abraham, Hillary; Chaiwoo Lee; Samantha Brady; Craig Fitzgerald; Bruce Mehler; Bryan Reimer; Joseph F. Coughlin (2017). "Autonomous Vehicles and Alternatives to Driving: Trust, Preferences, and Effects of Age". In: *96th Annual Meeting of the Transportation Research Board*.

Albert, Marc; Claudio Ruch; Emilio Frazzoli (2019). "Imbalance in Mobility-on-Demand Systems". In: *ACM Transactions on Spatial Algorithms and Systems* 5.2, pp. 1–22. ISSN: 23740353. DOI: 10.1145/3325914.

Alonso-González, María J.; Oded Cats; Niels van Oort; Sascha Hoogendoorn-Lanser; Serge Hoogendoorn (2020). "What are the determinants of the willingness to share rides in pooled on-demand services?" In: *Transportation* 1986.1, p. 98. ISSN: 0049-4488. DOI: 10.1007/s11116-020-10110-2.

Alonso-Mora, Javier; Samitha Samaranayake; Alex Wallar; Emilio Frazzoli; Daniela Rus (2017). "On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment". In: *Proceedings of the National Academy of Sciences*, pp. 462–467. DOI: 10.1073/pnas.1611675114.

Alonso-Mora, Javier; Alex Wallar; Daniela Rus (2017). "Predictive routing for autonomous mobility-on-demand systems with ride-sharing". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3583–3590. DOI: 10.1109/IROS.2017.8206203.

Arbib, James; Tony Seba (2017). *Rethinking Transportation 2020-2030*. URL: https://www.rethinkX.com/transportation (visited on 12/27/2019).

Atac, Selin; Nikola Obrenovic; Michel Bierlaire (2019). "A Holistic Decision Making Framework for a Vehicle Sharing System". In: *New Trends in Databases and Information Systems*. Vol. 1064. Communications in Computer and Information Science. Cham: Springer International Publishing, pp. 306–314. ISBN: 978-3-030-30277-1. DOI: 10.1007/978-3-030-30278-8_32.

Atasoy, Bilge; Takuro Ikeda; Xiang Song; Moshe E. Ben-Akiva (2015). "The concept and impact analysis of a flexible mobility on demand system". In: *Transportation Research Part C: Emerging Technologies* 56, pp. 373–392. ISSN: 0968090X. DOI: 10.1016/j.trc.2015.04.009.

Azevedo, Carlos Lima; Katarzyna Marczuk; Sebastián Raveau; Harold Soh; Muhammad Adnan; Kakali Basak; Harish Loganathan; Neeraj Deshmunkh; Der-Horng Lee; Emilio Frazzoli; Moshe Ben-Akiva (2016). "Microsimulation of Demand and Supply of Autonomous Mobility On Demand". In: *Transportation Research Record: Journal of the Transportation Research Board* 2564.1, pp. 21–30. ISSN: 0361-1981. DOI: 10.3141/2564-03.

Baker, Barrie M.; M. A. Ayechew (2003). "A genetic algorithm for the vehicle routing problem". In: *Computers & Operations Research* 30.5, pp. 787–800. ISSN: 03050548. DOI: 10.1016/S0305-0548(02)00051-5.

Bansal, Prateek; Kara M. Kockelman (2017). "Forecasting Americans' long-term adoption of connected and autonomous vehicle technologies". In: *Transportation Research Part A: Policy and Practice* 95, pp. 49–63. ISSN: 09658564. DOI: 10.1016/j.tra.2016.10.013.

Bansal, Prateek; Akanksha Sinha; Rubal Dua; Ricardo A. Daziano (2020). "Eliciting preferences of TNC users and drivers: Evidence from the United States". In: *Travel Behaviour and Society* 20, pp. 225–236. ISSN: 2214367X. DOI: 10.1016/j.tbs.2020.04.002.

Basu, Rounaq; Andrea Araldo; Arun Prakash Akkinepally; Bat Hen Nahmias Biran; Kalaki Basak; Ravi Seshadri; Neeraj Deshmukh; Nishant Kumar; Carlos Lima Azevedo; Moshe Ben-Akiva (2018). "Automated Mobility-on-Demand vs. Mass Transit: A Multi-Modal Activity-Driven Agent-Based Simulation Approach". In: *Transportation Research Record: Journal of the Transportation Research Board*, p. 036119811875863. ISSN: 0361-1981. DOI: 10.1177/0361198118758630.

Bauer, Gordon S.; Jeffery B. Greenblatt; Brian F. Gerke (2018). "Cost, Energy, and Environmental Impact of Automated Electric Taxi Fleets in Manhattan". In: *Environmental science & technology* 52.8, pp. 4920–4928. DOI: 10.1021/acs.est.7b04732.

Becker, Henrik; Milos Balac; Francesco Ciari; Kay W. Axhausen (2020). "Assessing the welfare impacts of Shared Mobility and Mobility as a Service (MaaS)". In: *Transportation Research Part A: Policy and Practice* 131, pp. 228–243. ISSN: 09658564. DOI: 10.1016/j.tra.2019.09.027.

Beer, Rachel; Candace Brakewood; Subrina Rahman; Jennifer Viscardi (2017). "Qualitative Analysis of Ride-Hailing Regulations in Major American Cities". In: *Transportation Research Record: Journal of the Transportation Research Board* 2650.1, pp. 84–91. ISSN: 0361-1981. DOI: 10.3141/2650-10.

Beheshtian, Arash; R. Richard Geddes; Omid M. Rouhani; Kara M. Kockelman; Axel Ockenfels; Peter Cramton; Wooseok Do (2020). "Bringing the efficiency of electricity market mechanisms to multimodal mobility across congested transportation systems". In: *Transportation Research Part A: Policy and Practice* 131, pp. 58–69. ISSN: 09658564. DOI: 10.1016/j.tra.2019.09.021.

Bellon, Tina (Sept. 2019). *Uber to limit drivers' app access to comply with NYC regulation*. URL: https://www.businessinsider.com/uber-to-limit-drivers-app-access-to-comply-with-nyc-regulation-2019-9 (visited on 12/01/2020).

Ben-Akiva, Moshe; Michel Bierlaire (1999). "Discrete Choice Methods and their Applications to Short Term Travel Decisions". In: *Handbook of Transportation Science*. Ed. by Frederick S. Hillier; Randolph W. Hall. Vol. 23. International Series in Operations Research & Management Science. Boston, MA: Springer US, pp. 5–33. ISBN: 978-1-4613-7370-4. DOI: 10.1007/978-1-4615-5203-1_2.

Bertsekas, Dimitri P. (2005). *Dynamic programming and optimal control*. 3. ed. Vol. 3. Athena scientific optimization and computation series. Belmont, Mass.: Athena Scientific. ISBN: 9781886529083.

Bertsimas, Dimitris; Patrick Jaillet; Sébastien Martin (2019). "Online Vehicle Routing: The Edge of Optimization in Large-Scale Applications". In: *Operations Research* 67.1, pp. 143–162. ISSN: 0030-364X. DOI: 10.1287/opre.2018.1763.

Bilali, Aledia; Florian Dandl; Ulrich Fastenrath; Klaus Bogenberger (2019a). "An Analytical Model for On-Demand Ride Sharing to Evaluate the Impact of Reservation, Detour and Maximum Waiting Time". In: *22nd IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 1715–1720. ISBN: 978-1-5386-7024-8. DOI: 10.1109/ITSC.2019.8917280.

Bilali, Aledia; Florian Dandl; Ulrich Fastenrath; Klaus Bogenberger (2019b). "Impact of service quality factors on ride sharing in urban areas". In: *6th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 1–8. DOI: 10.1109/MTITS.2019.8883364.

Bilali, Aledia; Roman Engelhardt; Florian Dandl; Ulrich Fastenrath; Klaus Bogenberger (2020). "Analytical and Agent-Based Model to Evaluate Ride-Pooling Impact Factors". In: *Transportation Research Record: Journal of the Transportation Research Board*, p. 036119812091766. ISSN: 0361-1981. DOI: 10.1177/0361198120917666.

Bimpikis, Kostas; Ozan Candogan; Daniela Saban (2019). "Spatial Pricing in Ride-Sharing Networks". In: *Operations Research* 67.3, pp. 744–769. ISSN: 0030-364X. DOI: 10.1287/opre.2018.1800.

Bischoff, Joschka; Ihab Kaddoura; Michal Maciejewski; Kai Nagel (2018). "Simulation-based optimization of service areas for pooled ride-hailing operators". In: *Procedia Computer Science* 130, pp. 816–823. ISSN: 18770509. DOI: 10.1016/j.procs.2018.04.069.

Bischoff, Joschka; Michal Maciejewski (2016a). "Autonomous Taxicabs in Berlin – A Spatiotemporal Analysis of Service Performance". In: *Transportation Research Procedia* 19.4, pp. 176–186. ISSN: 23521465. DOI: 10.1016/j.trpro.2016.12.078.

Bischoff, Joschka; Michal Maciejewski (2016b). "Simulation of city-wide replacement of private cars with autonomous taxis in Berlin". In: *Ambient Systems, Networks and Technologies* 83, pp. 237–244. DOI: 10.1016/j.procs.2016.04.121.

Boesch, Patrick M.; Francesco Ciari; Kay W. Axhausen (2016). "Autonomous Vehicle Fleet Sizes Required to Serve Different Levels of Demand". In: *Transportation Research Record: Journal of the Transportation Research Board* 2542.1, pp. 111–119. ISSN: 0361-1981. DOI: 10.3141/2542-13.

Boewing, Felix; Maximilian Schiffer; Mauro Salazar; Marco Pavone (2020). "A Vehicle Coordination and Charge Scheduling Algorithm for Electric Autonomous Mobility-on-Demand Systems". In: *American Control Conference (ACC)*. IEEE, pp. 248–255. ISBN: 978-1-5386-8266-1. DOI: 10.23919/ACC45564.2020.9147734.

Bogenberger, Klaus; Philipp Blum; Florian Dandl; Lisa-Sophie Hamm; Allister Loder; Patrick Malcolm; Martin Margreiter; Natalie Sautter (2021). *MobilityCoins – A new currency for the multimodal urban transportation system*. URL: http://arxiv.org/pdf/2107.13441v2.

Bongiovanni, Claudia; Mor Kaspi; Nikolas Geroliminis (2019). "The electric autonomous dial-a-ride problem". In: *Transportation Research Part B: Methodological* 122, pp. 436–456. ISSN: 01912615. DOI: 10.1016/j.trb.2019.03.004.

Bortolomiol, Stefano; Virginie Lurkin; Michel Bierlaire (2021). "Price-based regulation of oligopolistic markets under discrete choice models of demand". In: *Transportation* 59.8, p. 1817. ISSN: 0049-4488. DOI: 10.1007/s11116-021-10217-0.

Bösch, Patrick M.; Felix Becker; Henrik Becker; Kay W. Axhausen (2017). "Cost-based analysis of autonomous mobility services". In: *Transport Policy*. ISSN: 0967070X. DOI: 10.1016/j.tranpol.2017.09.005.

Bracher, Benedikt; Klaus Bogenberger (2017). "A dynamic prizing scheme for a congestion charging zone based on a network fundamental diagram". In: *5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 669–674. ISBN: 978-1-5090-6484-7. DOI: 10.1109/MTITS.2017.8005597.

Braekers, Kris; An Caris; Gerrit K. Janssens (2014). "Exact and meta-heuristic approach for a general heterogeneous dial-a-ride problem with multiple depots". In: *Transportation Research Part B: Methodological* 67, pp. 166–186. ISSN: 01912615. DOI: 10.1016/j.trb.2014.05.007.

Burns, Lawrence D.; William C. Jordan; Bonnie A. Scarborough (2013). "Transforming personal mobility". In: *The Earth Institute*.

Caiati, Valeria; Soora Rasouli; Harry Timmermans (2019). "Bundling, pricing schemes and extra features preferences for mobility as a service: Sequential portfolio choice experiment". In: *Transportation Research Part A: Policy and Practice*. ISSN: 09658564. DOI: 10.1016/j.tra.2019.09.029.

Campbell, Alexia Fernández (Sept. 2019). *California passes AB5 bill to regulate Uber, Lyft and the gig economy*. URL: https://www.vox.com/2019/9/11/20850878/california-passes-ab5-bill-uber-lyft (visited on 12/01/2020).

Carron, Andrea; Francesco Seccamonte; Claudio Ruch; Emilio Frazzoli; Melanie N. Zeilinger (2019). "Scalable Model Predictive Control for Autonomous Mobility-on-Demand Systems". In: *IEEE Transactions on Control Systems Technology*, pp. 1–10. ISSN: 1063-6536. DOI: 10.1109/TCST.2019.2954520.

Cetin, Tamer; Elizabeth Deakin (2019). "Regulation of taxis and the rise of ridesharing". In: *Transport Policy* 76, pp. 149–158. ISSN: 0967070X. DOI: 10.1016/j.tranpol.2017.09.002.

Charkhgard, Hadi; Mahdi Takalloo; Zulqarnain Haider (2020). "Bi-objective autonomous vehicle repositioning problem with travel time uncertainty". In: *4OR* 27.4, p. 735. ISSN: 1619-4500. DOI: 10.1007/s10288-019-00429-7.

Chen, Le; Alan Mislove; Christo Wilson (2015). "Peeking Beneath the Hood of Uber". In: *the 2015 ACM Conference*. Ed. by Kenjiro Cho; Kensuke Fukuda; Vivek Pai; Neil Spring, pp. 495–508. DOI: 10.1145/2815675.2815681.

Chen, T. Donna; Kara M. Kockelman (2016). "Operations of a shared, autonomous, electric vehicle fleet: implications of vehicle & charging infrastructure decisions". In: *Transportation Research Part A: Policy and Practice* 94, pp. 243–254. ISSN: 09658564. DOI: 10.1016/j.tra.2016.08.020.

Chiang, Wen-Chyuan; Robert A. Russell (1996). "Simulated annealing metaheuristics for the vehicle routing problem with time windows". In: *Annals of Operations Research* 63.1, pp. 3–27. ISSN: 0254-5330. DOI: 10.1007/BF02601637.

Clewlow, Regina R.; Gouri S. Mishra (2017). "Disruptive Transportation: The Adoption, Utilization, and Impacts of Ride-Hailing in the United States". In: *Institute of Transportation Studies, University of California, Davis* Research Research Report—UCD-ITS-RR-17-07.

Cohen, Tom; Clémence Cavoli (2019). "Automated vehicles: exploring possible consequences of government (non)intervention for congestion and accessibility". In: *Transport Reviews* 39.1, pp. 129–151. ISSN: 0144-1647. DOI: 10.1080/01441647.2018.1524401.

Conger, Kate (Nov. 2020). *Uber and Lyft Drivers in California Will Remain Contractors*. URL: https://www.nytimes.com/2020/11/04/technology/california-uber-lyft-prop-22.html (visited on 12/01/2020).

Cordeau, Jean-François; Gilbert Laporte (2007). "The dial-a-ride problem: models and algorithms". In: *Annals of Operations Research* 153.1, pp. 29–46. ISSN: 0254-5330. DOI: 10.1007/s10479-007-0170-8.

Correia, Gonçalo Homem de Almeida; Bart van Arem (2016). "Solving the User Optimum Privately Owned Automated Vehicles Assignment Problem (UO-POAVAP): A model to explore the impacts of self-driving vehicles on urban mobility". In: *Transportation Research Part B: Methodological* 87, pp. 64–88. ISSN: 01912615. DOI: 10.1016/j.trb.2016.03.002.

Daganzo, Carlos F.; Yanfeng Ouyang (2019). "A general model of demand-responsive transportation services: From taxi to ridesharing to dial-a-ride". In: *Transportation Research Part B: Methodological* 126, pp. 213–224. ISSN: 01912615. DOI: 10.1016/j.trb.2019.06.001.

Dandl, Florian; Klaus Bogenberger (2018). "Booking Processes in Autonomous Carsharing and Taxi Systems". In: *Proceedings of 7th Transport Research Arena, Vienna*. DOI: 10.5281/zenodo.1451436.

Dandl, Florian; Klaus Bogenberger (2019). "Comparing Future Autonomous Electric Taxis With an Existing Free-Floating Carsharing System". In: *IEEE Transactions on Intelligent Transportation Systems* 20.6, pp. 2037–2047. ISSN: 1524-9050. DOI: 10.1109/TITS.2018.2857208.

Dandl, Florian; Klaus Bogenberger; Hani S. Mahmassani (2019). "Autonomous Mobility-on-Demand Real-Time Gaming Framework". In: *6th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 1–10. ISBN: 978-1-5386-9484-8. DOI: 10.1109/MTITS.2019.8883286.

Dandl, Florian; Benedikt Bracher; Klaus Bogenberger (2017). "Microsimulation of an autonomous taxi-system in Munich". In: *5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 833–838. DOI: 10.1109/MTITS.2017.8005628.

Dandl, Florian; Roman Engelhardt; Michael Hyland; Gabriel Tilg; Klaus Bogenberger; Hani S. Mahmassani (2021). "Regulating mobility-on-demand services: Tri-level model and Bayesian optimization solution approach". In: *Transporta-*

tion Research Part C: Emerging Technologies 125.1, p. 103075. ISSN: 0968090X. DOI: 10.1016/j.trc.2021.103075.

Dandl, Florian; Fabian Fehn; Klaus Bogenberger; Fritz Busch (2020). "Pre-Day Scheduling of Charging Processes in Mobility-on-Demand Systems Considering Electricity Price and Vehicle Utilization Forecasts". In: *3rd Forum on Integrated and Sustainable Transportation Systems (FISTS)*. IEEE, pp. 127–134. ISBN: 978-1-7281-9503-2. DOI: 10.1109/FISTS46898.2020.9264862.

Dandl, Florian; Bernhard Grueber; Hanna Friese; Klaus Bogenberger (2019). "Design and Simulation of a Public-Transportation-Complimentary Autonomous Commuter Shuttle". In: *Transportation Research Procedia* 41, pp. 240–250. ISSN: 23521465. DOI: 10.1016/j.trpro.2019.09.043.

Dandl, Florian; Michael Hyland; Klaus Bogenberger; Hani S. Mahmassani (2019). "Evaluating the impact of spatio-temporal demand forecast aggregation on the operational performance of shared autonomous mobility fleets". In: *Transportation* 46.6, pp. 1975–1996. ISSN: 0049-4488. DOI: 10.1007/s11116-019-10007-9.

Dandl, Florian; Michael Hyland; Klaus Bogenberger; Hani S. Mahmassani (2020). "Dual-Horizon Forecasts and Repositioning Strategies for Operating Shared Autonomous Mobility Fleets". In: *99th Annual Meeting of the Transportation Research Board*.

Dandl, Florian; Gabriel Tilg; Majid Rostami-Shahrbabaki; Klaus Bogenberger (2020). "Network Fundamental Diagram Based Routing of Vehicle Fleets in Dynamic Traffic Simulations". In: *23rd IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 1–8. ISBN: 978-1-7281-4149-7. DOI: 10.1109/ITSC45102.2020.9294204.

Delling, Daniel; Andrew V. Goldberg; Thomas Pajor; Renato F. Werneck (2017). "Customizable Route Planning in Road Networks". In: *Transportation Science* 51.2, pp. 566–591. ISSN: 0041-1655. DOI: 10.1287/trsc.2014.0579.

Dijkstra, E. W. (1959). "A note on two problems in connexion with graphs". In: *Numerische Mathematik* 1.1, pp. 269–271. ISSN: 0029-599X. DOI: 10.1007/BF01386390.

Djavadian, Shadi; Joseph Y. J. Chow (2017). "Agent-based day-to-day adjustment process to evaluate dynamic flexible transport service policies". In: *Transportmetrica B: Transport Dynamics* 5.3, pp. 281–306. ISSN: 2168-0566. DOI: 10.1080/21680566.2016.1190674.

Dong, Xiaoxia; Matthew DiScenna; Erick Guerra (2019). "Transit user perceptions of driverless buses". In: *Transportation* 46.1, pp. 35–50. ISSN: 0049-4488. DOI: 10.1007/s11116-017-9786-y.

Duarte, Fábio; Carlo Ratti (2018). "The Impact of Autonomous Vehicles on Cities: A Review". In: *Journal of Urban Technology* 25.4, pp. 3–18. ISSN: 1063-0732. DOI: 10.1080/10630732.2018.1493883.

Engelhardt, Roman; Klaus Bogenberger (2021). "Benefits of Flexible Boarding Locations in On-Demand Ride-Pooling Systems". In: *7th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 1–6. ISBN: 978-1-7281-8995-6. DOI: 10.1109/MT-ITS49943.2021.9529284.

Engelhardt, Roman; Florian Dandl; Aledia Bilali; Klaus Bogenberger (2019). "Quantifying the Benefits of Autonomous On-Demand Ride-Pooling: A Simulation Study for Munich, Germany". In: *22nd IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 2992–2997. ISBN: 978-1-5386-7024-8. DOI: 10.1109/ITSC.2019.8916955.

Engelhardt, Roman; Florian Dandl; Klaus Bogenberger (2019). *Speed-up Heuristic for an On-Demand Ride-Pooling Algorithm*. URL: https://arxiv.org/abs/2007.14877.

Engelhardt, Roman; Patrick Malcolm; Florian Dandl; Klaus Bogenberger (2022). "Competition and Cooperation of Autonomous Ridepooling Services: Simulation of a Novel Broker Concept". In: *accepted to 101st Annual Meeting of the Transportation Research Board*.

Erdmann, Marvin; Florian Dandl; Klaus Bogenberger (2019). "Dynamic Car-Passenger Matching based on Tabu Search using Global Optimization with Time Windows". In: *8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO)*. Piscataway, NJ: IEEE, pp. 1–5. ISBN: 978-1-5386-7684-4. DOI: 10.1109/ICMSAO.2019.8880293.

Erdmann, Marvin; Florian Dandl; Klaus Bogenberger (2021). "Combining immediate customer responses and car–passenger reassignments in on-demand mobility services". In: *Transportation Research Part C: Emerging Technologies* 126.10, p. 103104. ISSN: 0278-3649. DOI: 10.1016/j.trc.2021.103104.

Erdmann, Marvin; Florian Dandl; Bernd Kaltenhaeuser; Klaus Bogenberger (2020). "Dynamic Car-Passenger Matching of Online and Reservation Requests". In: *99th Annual Meeting of the Transportation Research Board*.

Estandia, Alvaro; Maximilian Schiffer; Federico Rossi; Justin Luke; Emre Can Kara; Ram Rajagopal; Marco Pavone (2021). "On the Interaction between Autonomous Mobility on Demand Systems and Power Distribution Networks — An Optimal Power Flow Approach". In: *IEEE Transactions on Control of Network Systems*, p. 1. DOI: 10.1109/TCNS.2021.3059225.

Fagnant, Daniel J.; Kara M. Kockelman (2014). "The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios". In: *Transportation Research Part C: Emerging Technologies* 40, pp. 1–13. ISSN: 0968090X. DOI: 10.1016/j.trc.2013.12.001.

Fagnant, Daniel J.; Kara M. Kockelman (2018). "Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in Austin, Texas". In: *Transportation* 45.1, pp. 143–158. ISSN: 0049-4488. DOI: 10.1007/s11116-016-9729-z.

Fagnant, Daniel J.; Kara M. Kockelman; Prateek Bansal (2015). "Operations of Shared Autonomous Vehicle Fleet for Austin, Texas Market". In: *Transportation Research Record: Journal of the Transportation Research Board* 2536, pp. 98–106. ISSN: 0361-1981. DOI: 10.3141/2536-12.

Farhan, Javed; Donna Chen; Zhouyi Zhang (2018). "Leveraging Shared Autonmous Electric Vehicles For First/Last Mile Mobility". In: *97th Annual Meeting of the Transportation Research Board*.

Farhan, Javed; T. Donna Chen (2018). "Impact of ridesharing on operational efficiency of shared autonomous electric vehicle fleet". In: *Transportation Research Part C: Emerging Technologies* 93, pp. 310–321. ISSN: 0968090X. DOI: 10.1016/j.trc.2018.04.022.

Fehn, Fabian; Florian Noack; Fritz Busch (2019). "Modeling of Mobility On-Demand Fleet Operations Based on Dynamic Electricity Pricing". In: *6th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 1–6. ISBN: 978-1-5386-9484-8. DOI: 10.1109/MTITS.2019.8883370.

Feneri, Anna-Maria; Soora Rasouli; Harry J.P. Timmermans (2020). "Modeling the effect of Mobility-as-a-Service on mode choice decisions". In: *Transportation Letters* 12.2, pp. 1–8. ISSN: 1942-7867. DOI: 10.1080/19427867.2020.1730025.

Fiedler, Davide; Michal Certicky; Javier Alonso-Mora; Michal Cap (2018). "The Impact of Ridesharing in Mobility-on-Demand Systems: Simulation Case Study in Prague". In: *21st IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 1173–1178. ISBN: 978-1-7281-0321-1. DOI: 10.1109/ITSC.2018.8569451.

Fielbaum, Andres; Xiaoshan Bai; Javier Alonso-Mora (2021). "On-demand ridesharing with optimized pick-up and drop-off walking locations". In: *Transportation Research Part C: Emerging Technologies* 126.3, p. 103061. ISSN: 0968090X. DOI: 10.1016/j.trc.2021.103061.

Fielbaum, Andrés; Javier Alonso-Mora (2020). "Unreliability in ridesharing systems: Measuring changes in users' times due to new requests". In: *Transportation Research Part C: Emerging Technologies* 121, p. 102831. ISSN: 0968090X. DOI: 10.1016/j.trc.2020.102831.

Figliozzi, Miguel Andres; Hani S. Mahmassani; Patrick Jaillet (2007). "Pricing in Dynamic Vehicle Routing Problems". In: *Transportation Science* 41.3, pp. 302–318. ISSN: 0041-1655. DOI: 10.1287/trsc.1070.0193.

Filipovska, Monika; Hani S. Mahmassani (2020). "Reliable Least-Time Path Estimation and Computation in Stochastic Time-Varying Networks with Spatio-Temporal Dependencies". In: *23rd IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 1–6. ISBN: 978-1-7281-4149-7. DOI: 10.1109/ITSC45102.2020.9294650.

Fluri, Christian; Claudio Ruch; Julian Zilly; Jan Hakenberg; Emilio Frazzoli (2019). "Learning to Operate a Fleet of Cars". In: *22nd IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 2292–2298. ISBN: 978-1-5386-7024-8. DOI: 10.1109/ITSC.2019.8917533.

Fortin, Jacey (2017). *Does Uber Really Prevent Drunken Driving? It Depends on the Study.* URL: https://www.nytimes.com/2017/04/07/business/uber-drunk-driving-prevention.html (visited on 12/01/2020).

Fraedrich, Eva; Dirk Heinrichs; Francisco J. Bahamonde-Birke; Rita Cyganski (2019). "Autonomous driving, the built environment and policy implications". In: *Transportation Research Part A: Policy and Practice* 122.5, pp. 162–172. ISSN: 09658564. DOI: 10.1016/j.tra.2018.02.018.

Frei, Charlotte; Michael Hyland; Hani S. Mahmassani (2017). "Flexing service schedules: Assessing the potential for demand-adaptive hybrid transit via a stated preference approach". In: *Transportation Research Part C: Emerging Technologies* 76, pp. 71–89. ISSN: 0968090X. DOI: 10.1016/j.trc.2016.12.017.

Geisberger, Robert; Peter Sanders; Dominik Schultes; Daniel Delling (2008). "Contraction Hierarchies: Faster and Simpler Hierarchical Routing in Road Networks". In: *Experimental Algorithms*. Ed. by Catherine C. McGeoch. Vol. 5038. Lecture notes in computer science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 319–333. ISBN: 978-3-540-68548-7. DOI: 10.1007/978-3-540-68552-4_24.

Gendreau, Michel; Jean-Yves Potvin (2005). "Tabu Search". In: *Search Methodologies*. Ed. by Edmund K. Burke; Graham Kendall. Vol. 6. Boston, MA: Springer US, pp. 165–186. ISBN: 978-0-387-23460-1. DOI: 10.1007/0-387-28356-0_6.

Gkartzonikas, Christos; Konstantina Gkritza (2019). "What have we learned? A review of stated preference and choice studies on autonomous vehicles". In: *Transportation Research Part C: Emerging Technologies* 98, pp. 323–337. ISSN: 0968090X. DOI: 10.1016/j.trc.2018.12.003.

Greenblatt, Jeffery B.; Susan Shaheen (2015). "Automated Vehicles, On-Demand Mobility, and Environmental Impacts". In: *Current Sustainable/Renewable Energy Reports* 2.3, pp. 74–81. ISSN: 2196-3010. DOI: 10.1007/s40518-015-0038-5.

Gschwind, Timo; Stefan Irnich (2015). "Effective Handling of Dynamic Time Windows and Its Application to Solving the Dial-a-Ride Problem". In: *Transportation Science* 49.2, pp. 335–354. ISSN: 0041-1655. DOI: 10.1287/trsc.2014.0531.

Gurumurthy, Krishna M.; Kara M. Kockelman (2019). "Modeling Americans' Autonomous Vehicle Preferences: A Focus on Dynamic Ride-Sharing, Privacy, & Long-Distance Mode Choices". In: *98th Annual Meeting of the Transportation Research Board*.

Gutman, Ronald (2004). "Reach-Based Routing: A New Approach to Shortest Path Algorithms Optimized for Road Networks". In: *Proceedings 6th International Workshop on Algorithm Engineering and Experiments, SIAM*.

Hall, Jonathan; Cory Kendrick; Chris Nosko (2015). "The effects of Uber's surge pricing: A case study". In: *The University of Chicago Booth School of Business*.

Harb, Mustapha; Yu Xiao; Giovanni Circella; Patricia L. Mokhtarian; Joan L. Walker (2018). "Projecting travelers into a world of self-driving vehicles: estimating travel behavior implications via a naturalistic experiment". In: *Transportation* 45.6, pp. 1671–1685. ISSN: 0049-4488. DOI: 10.1007/s11116-018-9937-9.

Hardt, Cornelius; Klaus Bogenberger (2021). "Dynamic Pricing in Free-Floating Carsharing Systems - A Model Predictive Control Approach". In: *100th Annual Meeting of the Transportation Research Board*.

Hart, Peter; Nils Nilsson; Bertram Raphael (1968). "A Formal Basis for the Heuristic Determination of Minimum Cost Paths". In: *IEEE Transactions on Systems Science and Cybernetics* 4.2, pp. 100–107. ISSN: 0536-1567. DOI: 10.1109/TSSC.1968.300136.

He, Suining; Kang G. Shin (2019). "Spatio-temporal Adaptive Pricing for Balancing Mobility-on-Demand Networks". In: *ACM Transactions on Intelligent Systems and Technology* 10.4, pp. 1–28. ISSN: 21576904. DOI: 10.1145/3331450.

Henao, Alejandro; Wesley E. Marshall (2019a). "An analysis of the individual economics of ride-hailing drivers". In: *Transportation Research Part A: Policy and Practice* 130, pp. 440–451. ISSN: 09658564. DOI: 10.1016/j.tra.2019.09.056.

Henao, Alejandro; Wesley E. Marshall (2019b). "The impact of ride-hailing on vehicle miles traveled". In: *Transportation* 46.6, pp. 2173–2194. ISSN: 0049-4488. DOI: 10.1007/s11116-018-9923-2.

Herminghaus, Stephan (2019). "Mean field theory of demand responsive ride pooling systems". In: *Transportation Research Part A: Policy and Practice* 119, pp. 15–28. ISSN: 09658564. DOI: 10.1016/j.tra.2018.10.028.

Ho, Sin C.; W. Y. Szeto; Yong-Hong Kuo; Janny M.Y. Leung; Matthew Petering; Terence W.H. Tou (2018). "A survey of dial-a-ride problems: Literature review and recent developments". In: *Transportation Research Part B: Methodological* 111, pp. 395–421. ISSN: 01912615. DOI: 10.1016/j.trb.2018.02.001.

Hörl, S.; C. Ruch; F. Becker; E. Frazzoli; K. W. Axhausen (2019). "Fleet operational policies for automated mobility: A simulation assessment for Zurich". In: *Transportation Research Part C: Emerging Technologies* 102, pp. 20–31. ISSN: 0968090X. DOI: 10.1016/j.trc.2019.02.020.

Hörl, Sebastian (2017). "Agent-based simulation of autonomous taxi services with dynamic demand responses". In: *Procedia Computer Science* 109, pp. 899–904. ISSN: 18770509. DOI: 10.1016/j.procs.2017.05.418.

Hörl, Sebastian; Milos Balac; Kay W. Axhausen (2019). "Dynamic demand estimation for an AMoD system in Paris". In: *IEEE Intelligent Vehicle Symposium*. DOI: 10.1109/IVS.2019.8814051.

Horn, Mark E.T. (2002). "Fleet scheduling and dispatching for demand-responsive passenger services". In: *Transportation Research Part C: Emerging Technologies* 10.1, pp. 35–63. ISSN: 0968090X. DOI: 10.1016/S0968-090X(01)00003-1.

Hou, Yi; Venu Garikapati; Dustin Weigl; Alejandro Henao; Matthew Moniot; Joshua Sperling (2020). "Factors Influencing Willingness to Pool in Ride-Hailing Trips". In: *Transportation Research Record: Journal of the Transportation Research Board* 2674.5, pp. 419–429. ISSN: 0361-1981. DOI: 10.1177/0361198120915886.

Hu, Liang; Jing Dong (2020). "An Artificial-Neural-Network-Based Model for Real-Time Dispatching of Electric Autonomous Taxis". In: *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10. ISSN: 1524-9050. DOI: 10.1109/TITS.2020.3029141.

Hyland, Michael; Florian Dandl; Klaus Bogenberger; Hani S. Mahmassani (2019). "Integrating demand forecasts into the operational strategies of shared automated vehicle mobility services: spatial resolution impacts". In: *Transportation Letters*, pp. 1–6. ISSN: 1942-7867. DOI: 10.1080/19427867.2019.1691297.

Hyland, Michael; Hani S. Mahmassani (2017). "Taxonomy of Shared Autonomous Vehicle Fleet Management Problems to Inform Future Transportation Mobility". In:

Transportation Research Record: Journal of the Transportation Research Board 2653, pp. 26–34. ISSN: 0361-1981. DOI: 10.3141/2653-04.

Hyland, Michael; Hani S. Mahmassani (2018a). "Dynamic autonomous vehicle fleet operations: Optimization-based strategies to assign AVs to immediate traveler demand requests". In: *Transportation Research Part C: Emerging Technologies* 92, pp. 278–297. ISSN: 0968090X. DOI: 10.1016/j.trc.2018.05.003.

Hyland, Michael; Hani S. Mahmassani (2018b). "Sharing is Caring: Dynamic Autonomous Vehicle Fleet Operations under Demand Surges". In: *97th Annual Meeting of the Transportation Research Board*.

Hyland, Michael F.; Hani S. Mahmassani (2020). "Operational benefits and challenges of shared-ride automated mobility-on-demand services". In: *Transportation Research Part A: Policy and Practice* 134, pp. 251–270. ISSN: 09658564. DOI: 10.1016/j.tra.2020.02.017.

Hyytiä, Esa; Aleksi Penttinen; Reijo Sulonen (2012). "Non-myopic vehicle and route selection in dynamic DARP with travel time and workload objectives". In: *Computers & Operations Research* 39.12, pp. 3021–3030. ISSN: 03050548. DOI: 10.1016/j.cor.2012.03.002.

Iacobucci, Riccardo; Benjamin McLellan; Tetsuo Tezuka (2019). "Optimization of shared autonomous electric vehicles operations with charge scheduling and vehicle-to-grid". In: *Transportation Research Part C: Emerging Technologies* 100, pp. 34–52. ISSN: 0968090X. DOI: 10.1016/j.trc.2019.01.011.

Jaeger, Benedikt; Carsten Brickwedde; Markus Lienkamp (2018). "Multi-Agent Simulation of a Demand-Responsive Transit System Operated by Autonomous Vehicles". In: *Anual Meeting of the Transportation Research Board*.

Jiang, Weiwei; Lin Zhang (2018). "Evaluating the Effects of Double-Apping on the Smartphone-Based E-Hailing Service: A Simulation-Based Study". In: *IEEE Access* 6, pp. 6654–6667. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2797207.

Jing, Peng; Hanbin Hu; Fengping Zhan; Yuexia Chen; Yuji Shi (2020). "Agent-Based Simulation of Autonomous Vehicles: A Systematic Literature Review". In: *IEEE Access* 8, pp. 79089–79103. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.2990295.

Jittrapirom, Peraphan; Valeria Caiati; Anna-Maria Feneri; Shima Ebrahimigharehbaghi; María J. Alonso-González; Narayan Jishnu (2017). "Mobility as a Service: A Critical Review of Definitions, Assessments of Schemes, and Key Challenges". In: *Urban Planning* 2, pp. 13–15. DOI: 10.17645/up.v2i2.931.

Jung, Jaeyoung; Joseph Y.J. Chow; R. Jayakrishnan; Ji Young Park (2014). "Stochastic dynamic itinerary interception refueling location problem with queue delay for electric taxi charging stations". In: *Transportation Research Part C: Emerging Technologies* 40, pp. 123–142. ISSN: 0968090X. DOI: 10.1016/j.trc.2014.01.008.

Jung, Jaeyoung; R. Jayakrishnan (2014). "A Simulation Framework for Modeling Large-Scale Flexible Transit Systems". In: *Transportation Research Record: Journal of the Transportation Research Board* 2466, pp. 31–41. ISSN: 0361-1981. DOI: 10.3141/2466-04.

Jung, Jaeyoung; R. Jayakrishnan; Ji Young Park (2016). "Dynamic Shared-Taxi Dispatch Algorithm with Hybrid-Simulated Annealing". In: *Computer-Aided Civil and Infrastructure Engineering* 31.4, pp. 275–291. issn: 10939687. doi: 10.1111/mice.12157.

Kaltenhäuser, Bernd; Karl Werdich; Florian Dandl; Klaus Bogenberger (2020). "Market development of autonomous driving in Germany". In: *Transportation Research Part A: Policy and Practice* 132, pp. 882–910. issn: 09658564. doi: 10.1016/j.tra.2020.01.001.

Kamargianni, Maria; Melinda Matyas (2017). "The Business Ecosystem of Mobility-as-a-Service". In: *96th Annual Meeting of the Transportation Research Board*.

Kang, Namwoo; Fred M. Feinberg; Panos Y. Papalambros (2015). *Autonomous Electric Vehicle Sharing System Design*. New York, N.Y.: American Society of Mechanical Engineers. doi: 10.1115/DETC2015-46491.

Al-Kanj, Lina; Juliana Nascimento; Warren B. Powell (2020). "Approximate dynamic programming for planning a ride-hailing system using autonomous fleets of electric vehicles". In: *European Journal of Operational Research* 284.3, pp. 1088–1106. issn: 03772217. doi: 10.1016/j.ejor.2020.01.033.

Kleiner, Alexander; Bernhard Nebel; Vittorio Amos Ziparo (2011). "A Mechanism for Dynamic Ride Sharing Based on Parallel Auctions". In: *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, IJCAI-11*. Ed. by Toby Walsh. Menlo Park, Calif.: AAAI Press. isbn: 978-1-57735-515-1.

Kondor, Dániel; Paolo Santi; Diem-Trinh Le; Xiaohu Zhang; Adam Millard-Ball; Carlo Ratti (2020). "Addressing the "minimum parking" problem for on-demand mobility". In: *Scientific reports* 10.1, p. 15885. doi: 10.1038/s41598-020-71867-1.

Kozlov, M. K.; S. P. Tarasov; L. G. Khachiyan (1980). "The polynomial solvability of convex quadratic programming". In: *USSR Computational Mathematics and Mathematical Physics* 20.5, pp. 223–228. issn: 00415553. doi: 10.1016/0041-5553(80)90098-1.

Krueger, Rico; Taha H. Rashidi; John M. Rose (2016). "Preferences for shared autonomous vehicles". In: *Transportation Research Part C: Emerging Technologies* 69, pp. 343–355. issn: 0968090X. doi: 10.1016/j.trc.2016.06.015.

Kucharski, Rafał; Oded Cats (2020). "Exact matching of attractive shared rides (ExMAS) for system-wide strategic evaluations". In: *Transportation Research Part B: Methodological* 139, pp. 285–310. issn: 01912615. doi: 10.1016/j.trb.2020.06.006.

Lamotte, Raphaël; André de Palma; Nikolas Geroliminis (2017). "On the use of reservation-based autonomous vehicles for demand management". In: *Transportation Research Part B: Methodological* 99, pp. 205–227. issn: 01912615. doi: 10.1016/j.trb.2017.01.003.

Larsen, Allan (2000). "The dynamic vehicle routing problem". PhD Thesis. Technical University of Denmark.

Levin, Michael W. (2017). "Congestion-aware system optimal route choice for shared autonomous vehicles". In: *Transportation Research Part C: Emerging Technologies* 82, pp. 229–247. issn: 0968090X. doi: 10.1016/j.trc.2017.06.020.

Levin, Michael W.; Kara M. Kockelman; Stephen D. Boyles; Tianxin Li (2017). "A general framework for modeling shared autonomous vehicles with dynamic network-loading and dynamic ride-sharing application". In: *Computers, Environment and Urban Systems* 64, pp. 373–383. ISSN: 01989715. DOI: 10.1016/j.compenvurbsys.2017.04.006.

Li, Baoxiang; Dmitry Krushinsky; Tom van Woensel; Hajo A. Reijers (2016). "An adaptive large neighborhood search heuristic for the share-a-ride problem". In: *Computers & Operations Research* 66, pp. 170–180. ISSN: 03050548. DOI: 10.1016/j.cor.2015.08.008.

Li, Li; DianChao Lin; Theodoros Pantelidis; Joseph Chow; Saif Eddin Jabari (2019). "An Agent-based Simulation for Shared Automated Electric Vehicles with Vehicle Relocation*". In: *22nd IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 3308–3313. ISBN: 978-1-5386-7024-8. DOI: 10.1109/ITSC.2019.8917253.

Li, Li; Theodoros Pantelidis; Joseph Y.J. Chow; Saif Eddin Jabari (2021). "A real-time dispatching strategy for shared automated electric vehicles with performance guarantees". In: *Transportation Research Part E: Logistics and Transportation Review* 152, p. 102392. ISSN: 13665545. DOI: 10.1016/j.tre.2021.102392.

Li, Qing; Feixiong Liao (2020). "Incorporating vehicle self-relocations and traveler activity chains in a bi-level model of optimal deployment of shared autonomous vehicles". In: *Transportation Research Part B: Methodological* 140, pp. 151–175. ISSN: 01912615. DOI: 10.1016/j.trb.2020.08.001.

Li, Sen; Hamidreza Tavafoghi; Kameshwar Poolla; Pravin Varaiya (2019). "Regulating TNCs: Should Uber and Lyft set their own rules?" In: *Transportation Research Part B: Methodological* 129, pp. 193–225. ISSN: 01912615. DOI: 10.1016/j.trb.2019.09.008.

Liang, Xiao; Gonçalo Homem de Almeida Correia; Bart van Arem (2016). "Optimizing the service area and trip selection of an electric automated taxi system used for the last mile of train trips". In: *Transportation Research Part E: Logistics and Transportation Review* 93, pp. 115–129. ISSN: 13665545. DOI: 10.1016/j.tre.2016.05.006.

Liang, Xiao; Gonçalo Homem de Almeida Correia; Bart van Arem (2018). "Applying a model for trip assignment and dynamic routing of automated taxis with congestion: System performance in the cityof Delft, The Netherlands". In: *Transportation Research Record: Journal of the Transportation Research Board*. ISSN: 0361-1981. DOI: 10.1177/0361198118758048.

Liang, Xiao; Correia, Gonçalo Homem de Almeida; Kun An; Bart van Arem (2020). "Automated taxis' dial-a-ride problem with ride-sharing considering congestion-based dynamic travel times". In: *Transportation Research Part C: Emerging Technologies* 112, pp. 260–281. ISSN: 0968090X. DOI: 10.1016/j.trc.2020.01.024.

Lippoldt, Katrin; Tanja Niels; Klaus Bogenberger (2018). "Effectiveness of different incentive models in free-floating carsharing systems: A case study in Milan". In: *21st IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 1179–1185. ISBN: 978-1-7281-0321-1. DOI: 10.1109/ITSC.2018.8569242.

Lippoldt, Katrin; Tanja Niels; Klaus Bogenberger (2019). "Analyzing the Potential of User-Based Relocations on a Free-Floating Carsharing System in Cologne". In: *Transportation Research Procedia* 37, pp. 147–154. ISSN: 23521465. DOI: 10.1016/j.trpro.2018.12.177.

Litman, Todd (2019). *Autonomous vehicle implementation predictions: Implications for Transport Planning*. URL: https://www.vtpi.org/avip.pdf (visited on 12/27/2019).

Liu, Yang; Prateek Bansal; Ricardo Daziano; Samitha Samaranayake (2019). "A Framework to Integrate Mode Choice in the Design of Mobility-on-Demand Systems". In: *Transportation Research Part C: Emerging Technologies* 105, pp. 648–665. ISSN: 0968090X. DOI: 10.1016/j.trc.2018.09.022.

Liu, Yang; Samitha Samaranayake (2019). *Proactive rebalancing and speed-up techniques for on-demand high capacity vehicle pooling*. URL: http://arxiv.org/pdf/1902.03374v1.

Loeb, Benjamin; Kara M. Kockelman (2019). "Fleet performance and cost evaluation of a shared autonomous electric vehicle (SAEV) fleet: A case study for Austin, Texas". In: *Transportation Research Part A: Policy and Practice* 121, pp. 374–385. ISSN: 09658564. DOI: 10.1016/j.tra.2019.01.025.

Loeb, Benjamin; Kara M. Kockelman; Jun Liu (2018). "Shared autonomous electric vehicle (SAEV) operations across the Austin, Texas network with charging infrastructure decisions". In: *Transportation Research Part C: Emerging Technologies* 89, pp. 222–233. ISSN: 0968090X. DOI: 10.1016/j.trc.2018.01.019.

Ma, Tai-yu; Saeid Rasulkhani; Joseph Y.J. Chow; Sylvain Klein (2019). "A dynamic ridesharing dispatch and idle vehicle repositioning strategy with integrated transit transfers". In: *Transportation Research Part E: Logistics and Transportation Review* 128, pp. 417–442. ISSN: 13665545. DOI: 10.1016/j.tre.2019.07.002.

Maciejewski, Michal; Joschka Bischoff (2017). "Congestion effects of autonomous taxi fleets". In: *Transport* 45.9, pp. 1–10. ISSN: 1648-4142. DOI: 10.3846/16484142.2017.1347827.

Maciejewski, Michal; Joschka Bischoff; Kai Nagel (2016). "An Assignment-Based Approach to Efficient Real-Time City-Scale Taxi Dispatching". In: *IEEE Intelligent Systems* 31.1, pp. 68–77. ISSN: 1541-1672. DOI: 10.1109/MIS.2016.2.

Maget, Christoph; Juliane Pillat; Volker Wassmuth (2019). "Transport demand model for the Free State of Bavaria – basis for local transport planning". In: *Transportation Research Procedia* 41, pp. 219–228. ISSN: 23521465. DOI: 10.1016/j.trpro.2019.09.040.

Marczuk, Katarzyna Anna; Harold Soh Soon Hong; Carlos Miguel Lima Azevedo; Muhammad Adnan; Scott Drew Pendleton; Emilio Frazzoli; Der Horng Lee (2015). "Autonomous mobility on demand in SimMobility: Case study of the central business district in Singapore". In: *7th International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, pp. 167–172. DOI: 10.1109/ICCIS.2015.7274567.

Martinez, Luis M.; Gonçalo H. A. Correia; José M. Viegas (2015). "An agent-based simulation model to assess the impacts of introducing a shared-taxi system:

An application to Lisbon (Portugal)". In: *Journal of Advanced Transportation* 49.3, pp. 475–495. ISSN: 01976729. DOI: 10.1002/atr.1283.

Maurer, Markus; J Christian Gerdes; Barbara Lenz; Hermann Winner (2016). *Autonomous driving: technical, legal and social aspects*. Springer Nature.

Meyer, Gereon; Sven Beiker, eds. (2014). *Road Vehicle Automation*. Lecture Notes in Mobility. Cham, s.l.: Springer International Publishing. ISBN: 978-3-319-05989-1. DOI: 10.1007/978-3-319-05990-7.

Meyer, Gereon; Sven Beiker (2019). *Road Vehicle Automation 6*. Cham: Springer International Publishing. ISBN: 978-3-030-22932-0. DOI: 10.1007/978-3-030-22933-7.

Molenbruch, Yves; Kris Braekers; An Caris (2017). "Typology and literature review for dial-a-ride problems". In: *Annals of Operations Research* 259.1-2, pp. 295–325. ISSN: 0254-5330. DOI: 10.1007/s10479-017-2525-0.

Moody, Joanna; Jinhua Zhao (2020). "Adoption of Exclusive and Pooled TNC Services in Singapore and the US". In: *Journal of Transportation Engineering, Part A: Systems* 146.9, p. 04020102. ISSN: 2473-2907. DOI: 10.1061/JTEPBS.0000438.

Nahmias-Biran, Bat-hen; Jimi B. Oke; Nishant Kumar; Carlos Lima Azevedo; Moshe Ben-Akiva (2020). "Evaluating the impacts of shared automated mobility on-demand services: an activity-based accessibility approach". In: *Transportation*. ISSN: 0049-4488. DOI: 10.1007/s11116-020-10106-y.

Nair, Gopindra S.; Chandra R. Bhat (2021). "Sharing the road with autonomous vehicles: Perceived safety and regulatory preferences". In: *Transportation Research Part C: Emerging Technologies* 122, p. 102885. ISSN: 0968090X. DOI: 10.1016/j.trc.2020.102885.

Narayanan, Santhanakrishnan; Emmanouil Chaniotakis; Constantinos Antoniou (2020a). "Factors affecting traffic flow efficiency implications of connected and autonomous vehicles: A review and policy recommendations". In: *Policy Implications of Autonomous Vehicles*. Vol. 5. Advances in Transport Policy and Planning. Elsevier, pp. 1–50. ISBN: 9780128201916. DOI: 10.1016/bs.atpp.2020.02.004.

Narayanan, Santhanakrishnan; Emmanouil Chaniotakis; Constantinos Antoniou (2020b). "Shared autonomous vehicle services: A comprehensive review". In: *Transportation Research Part C: Emerging Technologies* 111, pp. 255–293. ISSN: 0968090X. DOI: 10.1016/j.trc.2019.12.008.

Navjyoth Sarma, J. S.; Daisik Nam; Michael F. Hyland; Felipe de Souza; Dingtong Yang; Arash Ghaffar; I. Omer Verbas (2020). "Effective and Efficient Fleet Dispatching Strategies for Dynamically Matching AVs to Travelers in Large-scale Transportation Systems". In: *23rd IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 1–6. ISBN: 978-1-7281-4149-7. DOI: 10.1109/ITSC45102.2020.9294340.

Niels, Tanja; Klaus Bogenberger (2017). "Booking Behavior of Free-Floating Carsharing Users". In: *Transportation Research Record: Journal of the Transportation Research Board* 2650.1, pp. 123–132. ISSN: 0361-1981. DOI: 10.3141/2650-15.

Nourinejad, Mehdi; Mohsen Ramezani (2019). "Ride-Sourcing Modeling and Pricing in Non-Equilibrium Two-Sided Markets". In: *Transportation Research Procedia* 38, pp. 833–852. ISSN: 23521465. DOI: 10.1016/j.trpro.2019.05.043.

Oke, Jimi B.; Youssef M. Aboutaleb; Arun Akkinepally; Carlos Lima Azevedo; Yafei Han; P. Christopher Zegras; Joseph Ferreira; Moshe E. Ben-Akiva (2019). "A novel global urban typology framework for sustainable mobility futures". In: *Environmental Research Letters* 14.9, p. 095006. DOI: 10.1088/1748-9326/ab22c7.

Oke, Jimi B.; Arun Prakash Akkinepally; Siyu Chen; Yifei Xie; Youssef M. Aboutaleb; Carlos Lima Azevedo; P. Christopher Zegras; Joseph Ferreira; Moshe Ben-Akiva (2020). "Evaluating the systemic effects of automated mobility-on-demand services via large-scale agent-based simulation of auto-dependent prototype cities". In: *Transportation Research Part A: Policy and Practice* 140, pp. 98–126. ISSN: 09658564. DOI: 10.1016/j.tra.2020.06.013.

Pardalos, Panos M.; Stephen A. Vavasis (1991). "Quadratic programming with one negative eigenvalue is NP-hard". In: *Journal of Global Optimization* 1.1, pp. 15–22. ISSN: 0925-5001. DOI: 10.1007/bf00120662.

Pavone, M.; S. L. Smith; E. Frazzoli; D. Rus (2012). "Robotic load balancing for mobility-on-demand systems". In: *The International Journal of Robotics Research* 31.7, pp. 839–854. ISSN: 0278-3649. DOI: 10.1177/0278364912444766.

Pavone, Marco (2015). "Autonomous Mobility-on-Demand Systems for Future Urban Mobility". In: *Autonomes Fahren*. Ed. by Markus Maurer; J. Christian Gerdes; Barbara Lenz; Hermann Winner. Vol. 58. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 399–416. ISBN: 978-3-662-45853-2. DOI: 10.1007/978-3-662-45854-9_19.

Peeta, Srinivas; Hani S. Mahmassani (1995). "System optimal and user equilibrium time-dependent traffic assignment in congested networks". In: *Annals of Operations Research* 60.1, pp. 81–113. ISSN: 0254-5330. DOI: 10.1007/BF02031941.

Pillac, Victor; Michel Gendreau; Christelle Guéret; Andrés L. Medaglia (2013). "A review of dynamic vehicle routing problems". In: *European Journal of Operational Research* 225.1, pp. 1–11. ISSN: 03772217. DOI: 10.1016/j.ejor.2012.08.015.

Pinto, Helen K.R.F.; Michael F. Hyland; Hani S. Mahmassani; İ. Ömer Verbas (2019). "Joint Design of Multimodal Transit Networks and Shared Autonomous Mobility Fleets". In: *Transportation Research Procedia* 38, pp. 98–118. ISSN: 23521465. DOI: 10.1016/j.trpro.2019.05.007.

Powell, Warren B. (2011). *Approximate dynamic programming: Solving the curses of dimensionality*. Second edition. Wiley series in probability and statistics. Hoboken, New Jersey: Wiley. ISBN: 9781118029169.

Psaraftis, Harilaos N.; Min Wen; Christos A. Kontovas (2016). "Dynamic vehicle routing problems: Three decades and counting". In: *Networks* 67.1, pp. 3–31. ISSN: 00283045. DOI: 10.1002/net.21628.

Rayle, Lisa; Danielle Dai; Nelson Chan; Robert Cervero; Susan Shaheen (2016). "Just a better taxi? A survey-based comparison of taxis, transit, and ridesourc-

ing services in San Francisco". In: *Transport Policy* 45, pp. 168–178. ISSN: 0967070X. DOI: 10.1016/j.tranpol.2015.10.004.

Regan, Amelia; Hani S. Mahmassani; Patrick Jaillet (1996). "Dynamic Decision Making for Commercial Fleet Operations Using Real-Time Information". In: *Transportation Research Record: Journal of the Transportation Research Board* 1537, pp. 91–97. ISSN: 0361-1981. DOI: 10.3141/1537-13.

Ropke, Stefan; David Pisinger (2006). "An Adaptive Large Neighborhood Search Heuristic for the Pickup and Delivery Problem with Time Windows". In: *Transportation Science* 40.4, pp. 455–472. ISSN: 0041-1655. DOI: 10.1287/trsc.1050.0135.

Rossi, Federico; Rick Zhang; Yousef Hindy; Marco Pavone (2018). "Routing autonomous vehicles in congested transportation networks: structural properties and coordination algorithms". In: *Autonomous Robots* 42.7, pp. 1427–1442. ISSN: 0929-5593. DOI: 10.1007/s10514-018-9750-5.

Ruch, Claudio; Sebastian Horl; Emilio Frazzoli (2018). "AMoDeus, a Simulation-Based Testbed for Autonomous Mobility-on-Demand Systems". In: *21st IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 3639–3644. DOI: 10.1109/ITSC.2018.8569961.

Ruch, Claudio; ChengQi Lu; Lukas Sieber; Emilio Frazzoli (2020). "Quantifying the Efficiency of Ride Sharing". In: *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–6. ISSN: 1524-9050. DOI: 10.1109/TITS.2020.2990202.

Salazar, Mauro; Federico Rossi; Maximilian Schiffer; Christopher H. Onder; Marco Pavone (2018). "On the Interaction between Autonomous Mobility-on-Demand and Public Transportation Systems". In: *21st IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 2262–2269. ISBN: 978-1-7281-0321-1. DOI: 10.1109/ITSC.2018.8569381.

Salazar, Mauro; Matthew Tsao; Izabel Aguiar; Maximilian Schiffer; Marco Pavone (2019). "A Congestion-aware Routing Scheme for Autonomous Mobility-on-Demand Systems". In: *18th European Control Conference (ECC)*. IEEE, pp. 3040–3046. ISBN: 978-3-907144-00-8. DOI: 10.23919/ECC.2019.8795897.

Santi, Paolo; Giovanni Resta; Michael Szell; Stanislav Sobolevsky; Steven H. Strogatz; Carlo Ratti (2014). "Quantifying the benefits of vehicle pooling with shareability networks". In: *Proceedings of the National Academy of Sciences of the United States of America* 111.37, pp. 13290–13294. ISSN: 0027-8424. DOI: 10.1073/pnas.1403657111.

Sayarshad, Hamid R.; Joseph Y. J. Chow (2016). "Survey and empirical evaluation of nonhomogeneous arrival process models with taxi data". In: *Journal of Advanced Transportation* 50.7, pp. 1275–1294. ISSN: 01976729. DOI: 10.1002/atr.1401.

Schilde, M.; K. F. Doerner; R. F. Hartl (2014). "Integrating stochastic time-dependent travel speed in solution methods for the dynamic dial-a-ride problem". In: *European Journal of Operational Research* 238.1, pp. 18–30. ISSN: 03772217. DOI: 10.1016/j.ejor.2014.03.005.

Séjournè, Thibault; Samitha Samaranayake; Siddhartha Banerjee (2018). "The Price of Fragmentation in Mobility-on-Demand Services". In: *Proceedings of*

the ACM on Measurement and Analysis of Computing Systems 2.2, pp. 1–26. ISSN: 2476-1249. DOI: 10.1145/3224425.

Shaheen, Susan (2018). "Shared Mobility: The Potential of Ridehailing and Pooling". In: *Three Revolutions*. Ed. by Daniel Sperling. Washington, DC: Island Press/Center for Resource Economics, pp. 55–76. ISBN: 978-1-61091-983-8. DOI: 10.5822/978-1-61091-906-7_3.

Shaheen, Susan; Nelson Chan (2016). "Mobility and the Sharing Economy: Potential to Facilitate the First- and Last-Mile Public Transit Connections". In: *Built Environment* 42.4, pp. 573–588. ISSN: 0263-7960. DOI: 10.2148/benv.42.4.573.

Shaheen, Susan A. Shaheen; Apaar Bansal Bansal; Nelson Chan Chan; Adam Cohen Cohen (2017). "Mobility and the sharing economy: industry developments and early understanding of impacts". In: *Low Carbon Mobility for Future Cities: Principles and applications*. Ed. by Hussein Dia Dia. Institution of Engineering and Technology, pp. 213–240. ISBN: 9781785611971. DOI: 10.1049/pbtr006e_ch10.

Shen, Yu; Hongmou Zhang; Jinhua Zhao (2018). "Integrating shared autonomous vehicle in public transportation system: A supply-side simulation of the first-mile service in Singapore". In: *Transportation Research Part A: Policy and Practice* 113, pp. 125–136. ISSN: 09658564. DOI: 10.1016/j.tra.2018.04.004.

Sheppard, Colin J. R.; Gordon S. Bauer; Brian F. Gerke; Jeffery B. Greenblatt; Alan T. Jenn; Anand R. Gopal (2019). "Joint Optimization Scheme for the Planning and Operations of Shared Autonomous Electric Vehicle Fleets Serving Mobility on Demand". In: *Transportation Research Record: Journal of the Transportation Research Board* 101.3, p. 036119811983827. ISSN: 0361-1981. DOI: 10.1177/0361198119838270.

Sieber, L.; C. Ruch; S. Hörl; K. W. Axhausen; E. Frazzoli (2020). "Improved public transportation in rural areas with self-driving cars: A study on the operation of Swiss train lines". In: *Transportation Research Part A: Policy and Practice* 134, pp. 35–51. ISSN: 09658564. DOI: 10.1016/j.tra.2020.01.020.

Simonetto, Andrea; Julien Monteil; Claudio Gambella (2019). "Real-time city-scale ridesharing via linear assignment problems". In: *Transportation Research Part C: Emerging Technologies* 101, pp. 208–232. ISSN: 0968090X. DOI: 10.1016/j.trc.2019.01.019.

Simoni, Michele D.; Kara M. Kockelman; Krishna M. Gurumurthy; Joschka Bischoff (2019). "Congestion pricing in a world of self-driving vehicles: An analysis of different strategies in alternative future scenarios". In: *Transportation Research Part C: Emerging Technologies* 98, pp. 167–185. ISSN: 0968090X. DOI: 10.1016/j.trc.2018.11.002.

Sobol, I. M. (1976). "Uniformly distributed sequences with an additional uniform property". In: *USSR Computational Mathematics and Mathematical Physics* 16.5, pp. 236–242. ISSN: 00415553. DOI: 10.1016/0041-5553(76)90154-3.

Spieser, Kevin; Samitha Samaranayake; Wolfgang Gruel; Emilio Frazzoli (2016). "Shared-Vehicle Mobility-on-Demand Systems: Fleet Operator's Guide to Rebalancing Empty Vehicles". In: *95th Annual Meeting of the Transportation Research Board*. Washington, D.C.

Srinivas, Niranjan; Andreas Krause; Sham M. Kakade; Matthias W. Seeger (2012). "Information-Theoretic Regret Bounds for Gaussian Process Optimization in the Bandit Setting". In: *IEEE Transactions on Information Theory* 58.5, pp. 3250–3265. ISSN: 0018-9448. DOI: 10.1109/TIT.2011.2182033.

Stiglic, Mitja; Niels Agatz; Martin Savelsbergh; Mirko Gradisar (2015). "The benefits of meeting points in ride-sharing systems". In: *Transportation Research Part B: Methodological* 82, pp. 36–53. ISSN: 01912615. DOI: 10.1016/j.trb.2015.07.025.

Stocker, Adam; Susan Shaheen (2017). *Shared automated vehicles: Review of business models*. eng. International Transport Forum Discussion Paper 2017-09. Paris. URL: http://hdl.handle.net/10419/194044.

Syed, Arslan Ali; Karim Akhnoukh; Bernd Kaltenhaeuser; Klaus Bogenberger (2019). "Neural Network Based Large Neighborhood Search Algorithm for Ride Hailing Services". In: *Progress in Artificial Intelligence*. Ed. by Paulo Moura Oliveira; Paulo Novais; Luís Paulo Reis. Vol. 11804. Lecture notes in computer science. Cham: Springer International Publishing, pp. 584–595. ISBN: 978-3-030-30240-5. DOI: 10.1007/978-3-030-30241-2_49.

Syed, Arslan Ali; Florian Dandl; Klaus Bogenberger (2021). "User-Assignment Strategy Considering Future Imbalance Impacts for Ride Hailing". In: *24th IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 2441–2446. ISBN: 978-1-7281-9142-3. DOI: 10.1109/ITSC48978.2021.9564559.

Syed, Arslan Ali; Florian Dandl; Bernd Kaltenhäuser; Klaus Bogenberger (2021). "Density Based Distribution Model for Repositioning Strategies of Ride Hailing Services". In: *Frontiers in Future Transportation* 2, p. 295. DOI: 10.3389/ffutr.2021.681451.

Syed, Arslan Ali; Bernd Kaltenhaeuser; Irina Gaponova; Klaus Bogenberger (2019). "Asynchronous Adaptive Large Neighborhood Search Algorithm for Dynamic Matching Problem in Ride Hailing Services". In: *22nd IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 3006–3012. ISBN: 978-1-5386-7024-8. DOI: 10.1109/ITSC.2019.8916943.

Tachet, R.; O. Sagarra; P. Santi; G. Resta; M. Szell; S. H. Strogatz; C. Ratti (2017). "Scaling Law of Urban Ride Sharing". In: *Scientific reports* 7, p. 42868. DOI: 10.1038/srep42868.

Talluri, Kalyan T.; Garrett J. van Ryzin (2004). *The theory and practice of revenue management*. International Series in Operations Research & Management Science. New York NY: Springer. ISBN: 978-0-387-27391-4.

Tirachini, Alejandro (2020). "Ride-hailing, travel behaviour and sustainable mobility: an international review". In: *Transportation* 47.4, pp. 2011–2047. ISSN: 0049-4488. DOI: 10.1007/s11116-019-10070-2.

Tirachini, Alejandro; Emmanouil Chaniotakis; Mohamed Abouelela; Constantinos Antoniou (2020). "The sustainability of shared mobility: Can a platform for shared rides reduce motorized traffic in cities?" In: *Transportation Research Part C: Emerging Technologies* 117, p. 102707. ISSN: 0968090X. DOI: 10.1016/j.trc.2020.102707.

Tirachini, Alejandro; Ricardo Hurtubia; Thijs Dekker; Ricardo A. Daziano (2017). "Estimation of crowding discomfort in public transport: Results from Santiago de Chile". In: *Transportation Research Part A: Policy and Practice* 103.3, pp. 311–326. ISSN: 09658564. DOI: 10.1016/j.tra.2017.06.008.

Turan, Berkay; Ramtin Pedarsani; Mahnoosh Alizadeh (2020). "Dynamic pricing and fleet management for electric autonomous mobility on demand systems". In: *Transportation Research Part C: Emerging Technologies* 121, p. 102829. ISSN: 0968090X. DOI: 10.1016/j.trc.2020.102829.

Vakayil; Akhil; Gruel; Wolfgang; Samaranayake; Samitha (2017). "Integrating Shared-Vehicle Mobility-on-Demand Systems with Public Transit". In: *96th Annual Meeting of the Transportation Research Board*.

Vazifeh, M. M.; P. Santi; G. Resta; S. H. Strogatz; C. Ratti (2018). "Addressing the minimum fleet problem in on-demand urban mobility". In: *Nature* 557.7706, pp. 534–538. DOI: 10.1038/s41586-018-0095-1.

Vidal, Thibaut; Teodor Gabriel Crainic; Michel Gendreau; Nadia Lahrichi; Walter Rei (2012). "A Hybrid Genetic Algorithm for Multidepot and Periodic Vehicle Routing Problems". In: *Operations Research* 60.3, pp. 611–624. ISSN: 0030-364X. DOI: 10.1287/opre.1120.1048.

Vosooghi, Reza; Jakob Puchinger; Joschka Bischoff; Marija Jankovic; Anthony Vouillon (2020). "Shared autonomous electric vehicle service performance: Assessing the impact of charging infrastructure". In: *Transportation Research Part D: Transport and Environment* 81, p. 102283. ISSN: 13619209. DOI: 10.1016/j.trd.2020.102283.

Vosooghi, Reza; Jakob Puchinger; Marija Jankovic; Anthony Vouillon (2019). "Shared autonomous vehicle simulation and service design". In: *Transportation Research Part C: Emerging Technologies* 107, pp. 15–33. ISSN: 0968090X. DOI: 10.1016/j.trc.2019.08.006.

Walker, Jonathan; Charlie Johnson (2016). *Peak Car Ownership: The Market Opportunity of Electric Automated Mobility Services*. URL: http://www.rmi.org/peak_car_ownership (visited on 12/27/2019).

Walker, Warren E.; Vincent A.W.J. Marchau (2017). "Dynamic adaptive policymaking for the sustainable city: The case of automated taxis". In: *International Journal of Transportation Science and Technology* 6.1, pp. 1–12. ISSN: 20460430. DOI: 10.1016/j.ijtst.2017.03.004.

Wallar, Alex; Wilko Schwarting; Javier Alonso-Mora; Daniela Rus (2019). "Optimizing Multi-class Fleet Compositions for Shared Mobility-as-a-Service". In: *22nd IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 2998–3005. ISBN: 978-1-5386-7024-8. DOI: 10.1109/ITSC.2019.8916904.

Wallar, Alex; Menno van der Zee; Javier Alonso-Mora; Daniela Rus (2018). "Vehicle Rebalancing for Mobility-on-Demand Systems with Ride-Sharing". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4539–4546. ISBN: 978-1-5386-8094-0. DOI: 10.1109/IROS.2018.8593743.

Wang, Senlei; Correia, Goncalo Homem de Almeida; Hai Xiang Lin (2019). "Exploring the Performance of Different On-Demand Transit Services Provided by a

Fleet of Shared Automated Vehicles: An Agent-Based Model". In: *Journal of Advanced Transportation* 2019, pp. 1–16. ISSN: 01976729. DOI: 10.1155/2019/7878042.

Wang, Xiaolei; Wei Liu; Hai Yang; Dan Wang; Jieping Ye (2019). "Customer behavioural modelling of order cancellation in coupled ride-sourcing and taxi markets". In: *Transportation Research Procedia* 38, pp. 853–873. ISSN: 23521465. DOI: 10.1016/j.trpro.2019.05.044.

Washbrook, Kevin; Wolfgang Haider; Mark Jaccard (2006). "Estimating commuter mode choice: A discrete choice analysis of the impact of road pricing and parking charges". In: *Transportation* 33.6, pp. 621–639. ISSN: 0049-4488. DOI: 10.1007/s11116-005-5711-x.

Weikl, Simone; Klaus Bogenberger (2013). "Relocation Strategies and Algorithms for Free-Floating Car Sharing Systems". In: *IEEE Intelligent Transportation Systems Magazine* 5.4, pp. 100–111. ISSN: 1939-1390. DOI: 10.1109/MITS.2013.2267810.

Weikl, Simone; Klaus Bogenberger (2015). "A practice-ready relocation model for free-floating carsharing systems with electric vehicles – Mesoscopic approach and field trial results". In: *Transportation Research Part C: Emerging Technologies* 57, pp. 206–223. ISSN: 0968090X. DOI: 10.1016/j.trc.2015.06.024.

Wen, Jian; Yu Xin Chen; Neema Nassir; Jinhua Zhao (2018). "Transit-oriented autonomous vehicle operation with integrated demand-supply interaction". In: *Transportation Research Part C: Emerging Technologies* 97, pp. 216–234. ISSN: 0968090X. DOI: 10.1016/j.trc.2018.10.018.

Wen, Jian; Neema Nassir; Jinhua Zhao (2019). "Value of demand information in autonomous mobility-on-demand systems". In: *Transportation Research Part A: Policy and Practice* 121, pp. 346–359. ISSN: 09658564. DOI: 10.1016/j.tra.2019.01.018.

Wen, Jian; Jinhua Zhao; Patrick Jaillet (2017). "Rebalancing shared mobility-on-demand systems: A reinforcement learning approach". In: *20th IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 220–225. DOI: 10.1109/ITSC.2017.8317908.

Wenzel, Tom; Clement Rames; Eleftheria Kontou; Alejandro Henao (2019). "Travel and energy implications of ridesourcing service in Austin, Texas". In: *Transportation Research Part D: Transport and Environment* 70, pp. 18–34. ISSN: 13619209. DOI: 10.1016/j.trd.2019.03.005.

Wilkes, Gabriel; Roman Engelhardt; Lars Briem; Florian Dandl; Peter Vortisch; Klaus Bogenberger; Martin Kagerbauer (2021). "Self-Regulating Demand and Supply Equilibrium in Joint Simulation of Travel Demand and a Ride-Pooling Service". In: *Transportation Research Record: Journal of the Transportation Research Board*, p. 036119812199714. ISSN: 0361-1981. DOI: 10.1177/0361198121997140.

Winter, Konstanze; Oded Cats; Gonçalo Correia; Bart van Arem (2018). "Performance analysis and fleet requirements of automated demand-responsive transport systems as an urban public transport service". In: *International Journal of Transportation Science and Technology* 7.2, pp. 151–167. ISSN: 20460430. DOI: 10.1016/j.ijtst.2018.04.004.

Winter, Konstanze; Oded Cats; Karel Martens; Bart van Arem (2020). "Relocating shared automated vehicles under parking constraints: assessing the impact

of different strategies for on-street parking". In: *Transportation* 114, p. 462. ISSN: 0049-4488. DOI: 10.1007/s11116-020-10116-w.

Winter, Konstanze; Oded Cats; Karel Martens; Bart van Arem (2021). "Parking space for shared automated vehicles: How less can be more". In: *Transportation Research Part A: Policy and Practice* 143, pp. 61–77. ISSN: 09658564. DOI: 10.1016/j.tra.2020.11.008.

Wittmann, Michael; Lorenz Neuner; Markus Lienkamp (2020). "A Predictive Fleet Management Strategy for On-Demand Mobility Services: A Case Study in Munich". In: *Electronics* 9.6, p. 1021. DOI: 10.3390/electronics9061021.

Xu, Xiang; Hani S. Mahmassani; Ying Chen (2019). "Privately Owned Autonomous Vehicle Optimization Model Development and Integration with Activity-Based Modeling and Dynamic Traffic Assignment Framework". In: *Transportation Research Record: Journal of the Transportation Research Board* 8.3, p. 036119811985207. ISSN: 0361-1981. DOI: 10.1177/0361198119852072.

Xu, Zhengtian; Yafeng Yin; Jieping Ye (2019). "On the supply curve of ride-hailing systems". In: *Transportation Research Procedia* 38, pp. 37–55. ISSN: 23521465. DOI: 10.1016/j.trpro.2019.05.004.

Yang, Hai; Min Ye; Wilson H. Tang; S. C. Wong (2005). "Regulating taxi services in the presence of congestion externality". In: *Transportation Research Part A: Policy and Practice* 39.1, pp. 17–40. ISSN: 09658564. DOI: 10.1016/j.tra.2004.05.004.

Yang, Jian; Patrick Jaillet; Hani S. Mahmassani (2004). "Real-Time Multivehicle Truckload Pickup and Delivery Problems". In: *Transportation Science* 38.2, pp. 135–148. ISSN: 0041-1655. DOI: 10.1287/trsc.1030.0068.

Yap, Menno D.; Gonçalo Correia; Bart van Arem (2016). "Preferences of travellers for using automated vehicles as last mile public transport of multimodal train trips". In: *Transportation Research Part A: Policy and Practice* 94, pp. 1–16. ISSN: 09658564. DOI: 10.1016/j.tra.2016.09.003.

Yu, James J. Q.; Albert Y. S. Lam; Zhiyi Lu (2018). "Double Auction-Based Pricing Mechanism for Autonomous Vehicle Public Transportation System". In: *IEEE Transactions on Intelligent Vehicles* 3.2, pp. 151–162. ISSN: 2379-8904. DOI: 10.1109/TIV.2018.2804161.

Yu, Jiangbo; Michael F. Hyland (2020). "A generalized diffusion model for preference and response time: Application to ordering mobility-on-demand services". In: *Transportation Research Part C: Emerging Technologies* 121, p. 102854. ISSN: 0968090X. DOI: 10.1016/j.trc.2020.102854.

Yurtsever, Ekim; Jacob Lambert; Alexander Carballo; Kazuya Takeda (2020). "A Survey of Autonomous Driving: Common Practices and Emerging Technologies". In: *IEEE Access* 8, pp. 58443–58469. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.2983149.

Zachariah, Jaison; Jingkang Gao; Alain Kornhauser; Talal Mufti (2014). "Uncongested Mobility for All: A Proposal for an Area Wide Autonomous Taxi System in New Jersey". In: *93rd Annual Meeting of the Transportation Research Board*.

Zgraggen, Jannik; Matthew Tsao; Mauro Salazar; Maximilian Schiffer; Marco Pavone (2019). "A Model Predictive Control Scheme for Intermodal Au-

tonomous Mobility-on-Demand". In: *22nd IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 1953–1960. ISBN: 978-1-5386-7024-8. DOI: 10.1109/ITSC.2019.8917521.

Zhang, Hongcai; Colin J. R. Sheppard; Timothy E. Lipman; Scott J. Moura (2019). "Joint Fleet Sizing and Charging System Planning for Autonomous Electric Vehicles". In: *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–14. ISSN: 1524-9050. DOI: 10.1109/TITS.2019.2946152.

Zhang, Kenan; Marco Nie (2019). "To Pool or Not to Pool: Equilibrium, Pricing and Regulation". In: *SSRN Electronic Journal* 114.3, p. 18. ISSN: 1556-5068. DOI: 10.2139/ssrn.3497808.

Zhang, R.; M. Pavone (2016). "Control of robotic mobility-on-demand systems: A queueing-theoretical perspective". In: *The International Journal of Robotics Research* 35.1-3, pp. 186–203. ISSN: 0278-3649. DOI: 10.1177/0278364915581863.

Zhang, Rick; Federico Rossi; Marco Pavone (2016). "Model predictive control of autonomous mobility-on-demand systems". In: *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1382–1389. DOI: 10.1109/ICRA.2016.7487272.

Zhang, Wenbo; Harsha Honnappa; Satish V. Ukkusuri (2019). "Modeling Urban Taxi Services with E-Hailings: A Queueing Network Approach". In: *Transportation Research Procedia* 38, pp. 751–771. ISSN: 23521465. DOI: 10.1016/j.trpro.2019.05.039.

Zhang, Wenwen; Subhrajit Guhathakurta; Jinqi Fang; Ge Zhang (2015). "Exploring the impact of shared autonomous vehicles on urban parking demand: An agent-based simulation approach". In: *Sustainable Cities and Society* 19, pp. 34–45. ISSN: 22106707. DOI: 10.1016/j.scs.2015.07.006.

Zhou, Meng; Donggen Wang (2019). "Generational differences in attitudes towards car, car ownership and car use in Beijing". In: *Transportation Research Part D: Transport and Environment* 72.2, pp. 261–278. ISSN: 13619209. DOI: 10.1016/j.trd.2019.05.008.

Zhu; Shirley; Kornhauser; Alain (2017). "Interplay Between Fleet Size, Level of Service, and Empty Vehicle Repositioning Strategies in Large-Scale, Shared-Ride Autonomous Taxi Mobility-on-Demand Scenarios". In: *96th Annual Meeting of the Transportation Research Board*.

Zwick, Felix; Kay W. Axhausen (2020). "Impact of Service Design on Urban Ride-pooling Systems". In: *23rd IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 1–6. ISBN: 978-1-7281-4149-7. DOI: 10.1109/ITSC45102.2020.9294289.

Zwick, Felix; Nico Kuehnel; Rolf Moeckel; Kay W. Axhausen (2021). "Agent-based simulation of city-wide autonomous ride-pooling and the impact on traffic noise". In: *Transportation Research Part D: Transport and Environment* 90, p. 102673. ISSN: 13619209. DOI: 10.1016/j.trd.2020.102673.

Appendix

A Mathematical Notation

A mathematical notation helps to formulate problems briefly and precisely. Following table summarizes the notation of variables and parameters and briefly describes their use. For clarity, parameters, which are only used and described within one certain paragraph, are not enlisted. Due to the amount of models required in the AMoD regulation model, some characters have to be overloaded. However, their meaning is clear from the context of the respective paragraphs.

Notation	Name	Further Description
A	action	typically state-dependent, i.e. $A(S(t))$
b	request bundle	a set of requests ($\in R$) that should be connected in a route
B	set of bundles	B_v represents the set of all request bundles that could be served by a vehicle v
c	cost coefficient	super- and subscripts are context specific; coefficients are utilized if costs can be separated into a constant (coefficient) and a (trip-dependent) characteristic value
C	cost function	super- and subscripts are context specific; utilized if coefficient notation is not sufficient to describe costs
d	distance	can be defined by, e.g., an edge $e \in E$, a route ϕ, a vehicle plan ξ, or a complete vehicle trajectory (sub-index $v \in V$); generally given in meters
D	price sensitivity	used to describe the price sensitivity of demand
E	network edges	refers to the set of all edges in a network graph G; an edge is defined by its start and end node (elements of N) and typically contains cost information (travel time, distance)
f	fare	travelers have to pay fares to PT and AMoD operators for traveling with their respective mobility services
F	control function	objective function of operator control action; super-script determines the action, e.g., R for request assignment
G	network graph	consists of nodes N and edges E denoted by $G = (N, E)$; sub-index can refer to type of network, e.g., PT or street network
h	reachability bandwidth	parameter that determines the radius of reachability for zone-to-zone correlations of forecasts
H	heaviside function	$H(x)$ is 0 for $x < 0$, 0.5 for $x = 0$ and 1 for $x > 0$
$\mathcal{H}_\zeta$	Bessel function	of order ζ
i	time step index	

Notation	Name	Further Description
I	zone imbalance	positive (negative) value reflects vehicle surplus (deficit); sub-index refers to zone and super-index to valid time horizon
j	forecast time step index	
k	density	macroscopic traffic flow measure for number of vehicles per meter
K	kernel density	kernel density distribution (used in different vector spaces based on context)
m_r	mode choice variable	mode choice of traveler r from set M
M	mode choice set	e.g., PV, PT, or AMoD
N	network nodes	refers to the set of all nodes in a network graph G
N^{rv}	RV-heuristic	refers to the number of vehicles selected by an RV-heuristic to limit the number of considered vehicles for the user-vehicle assignment processes
$\mathcal{N}(\mu, \sigma)$	normal distribution	with mean μ and standard deviation σ
od	OD-route	refers to the fastest route ϕ between two network points x^o and x^d
OD	OD-pair	refers to the set of all possible routes ϕ between two network points x^o and x^d; can be used to describe total demand between these points
P	profit	operator objective; measured in €
q	flow	macroscopic traffic flow measure for number of vehicles passing per second
R	requests / travelers	refers to the set of all requests, travelers or users; super-indices refer to certain subsets based on their state
s	exogenous state variables	time-dependent description of AMoD state variables that are independent of operator control
S	(AMoD) system state	time-dependent description of all AMoD state variables; contains vehicle positions, assigned plans, revealed demand, demand estimations, etc.; $S^P(t)$ denotes post-state at time t (after actions $A(t)$ are performed)
t	time	can be used in many contexts described by the respective sub- and super-indices; generally given in seconds
T	time horizon	can be used in many contexts described by the respective sub- and super-indices; generally given in seconds
U_r^m	travel utility	traveler objective; sub-index refers to a request $r \in R$ and super-index to a certain mode $m \in M$; generally given in €
V	vehicle fleet	set of all AMoD vehicles; super-indices refer to certain subsets depending on the state of a vehicle
W	social welfare	regulator objective; measured in €
x	network position	described by triple, where the first two entries refer to the start and end node of an edge ($e \in E$) and the third entry is the relative position on the edge, where 0 denotes the start and 1 the end of the edge; there are many contexts that describe the respective sub- and super-indices
X	stop	part of vehicle plan; $X = (x, t^s, t^l, t^e, R^+, R^-, l)$ is defined by a network position $x \in N$, a stop duration t_s, possibly a latest arrival time t^l and an earliest departure time t^e, the sets of boarding and disembarking requests denoted by R^+ and R^-, and a flag indicating locked stops
Z	Zones	set of all zones that build a disjoint union of the operating area

Notation	Name	Further Description
α	regulator decision variable	super- and subscripts are context specific
β	operator decision variable	super- and subscripts are context specific
γ	discount factor	used to weigh immediate rewards versus future rewards
Λ	stochastic demand information	typically refers to aggregated data on zone a level and a certain time horizon T; two sub-indices indicate OD information, one index indicates only departure information; $\bar{\Lambda}_z$ indicates a trip arrival estimation
δ^{ij}	selection tensor	Kronecker delta with two indices, see equation (3.28) for four-index definition
ζ	travel time factor	utilized in dynamic network to represent the ratio of current and free flow travel times on edges ($e \in E$)
Θ	state transition operator	transforms post-state of one time step to the state of the next time step; contains all vehicle movements, boarding and charging processes, as well as the revelation of new requests
ξ	vehicle plan	a sequence of stops X, where vehicles need to stop for several reasons, e.g., to wait for customers, let them board or disembark; these stops contain the time constraints that should be considered; the vehicle plan also defines the route ϕ through these stops (typically fastest route); sub-index v refers to a certain vehicle in V
Ξ	set of vehicle plans	typically refers to some set of feasible vehicle plans, super- and sub-indices are explained in the context
ρ	passengers/seats	ρ_v determines number of available seats of vehicle v, ρ_r number of passengers in request r
τ	time constraint	super-indices w, d for waiting time and in-vehicle driving time, respectively; generally measured in seconds
ϕ	route	a route is defined by a start position x_s, an end position x_e, and optionally a sequence of connected edges ($e \in E$) between them; if the sequence is not explicitly given, ϕ refers to the fastest route between x_s and x_e; in context of a request ϕ_r refers to the fastest route between a travelers origin and destination locations
ψ	expected reward (forecast)	used in repositioning formulation as expected rewarding for matching of future demand and supply
Ψ	reward for operator action	related to operator profit P of an action A
Ω	(operating) area	can be divided into a disjoint set of zones Z, where Ω_z refers to the area of zone z

Table A1: Mathematical notation used in this thesis in alphabetical order, where Latin letters are listed before Greek letter.

B Collection of Selected Algorithms

B.1 Computation of Number of Routes

Let $n_v = |R_v^o|$ be the number of on-board requests and $n_w = |R^w|$ the number of assigned requests waiting for pick-up. Then the number of possible routes to serve these requests N can be determined by following recursive algorithm:

Algorithm 1: get_nr_routes

```
Data: n_v ≥ 0; n_w ≥ 0;
Result: N = get_nr_routes(n_v, n_w)
n_D ← n_v + n_do ;                                   /* Number of drop-offs */
n_S ← n_v + 2 × n_do ;                               /* Number of stops */
if n_D ≤ 1 then
    if n_D = 0 then
        N ← 0;
    else
        N ← 1;
    end
else
    N ← 0;
    if n_w ≥ 1 then
        n_S^p ← n_S − 2 ;                            /* Stops without one waiting request */
        d ← 0 ;                                      /* new drop-off at any position d */
        while d ≤ n_s^p do
            N ← N + (d + 1)× get_nr_routes(n_o, n_w − 1) ;   /* d + 1 possibilities to include pick-up
              stop */
            d ← d + 1;
        end
    else
        n_S^p ← n_S − 1 ;                            /* Stops without one on-board request */
        N ← N + (n_S^p + 1)× get_nr_routes(n_o − 1, n_w) ;   /* n_S^p + 1 possibilities to include drop-off
          stop */
    end
end
```

B.2 Routing

Backwards-Directed Vehicle Search Algorithm

Let $G = (N, E)$ be a graph representing a street network and $C : E \to \mathbb{R}$ be the current travel times of each edge. Moreover, let V be the set of AMoD vehicles with current vehicle plans ξ_v for all $v \in V$ at time t. Moreover, let r be a new request asking for immediate pick-up at location x_r^p and τ_r^w the latest pick-up time.

Then the approach illustrated in Fig. 3.2 and described in Algorithm 2 is an efficient methodology to find all vehicles V_r that can serve this request. It consists of two parts: (i) the determination of general vehicle availability and (ii) a request-specific backwards directed one-to-many Dijkstra route search algorithm.

A few comments are noteworthy:
(i) The algorithm can also return the mappings $costs$ and $next$ as additional outputs if specific route costs and routes are required.
(ii) The implementation with priority queue (PQ) automatically takes care of sorting in new

nodes and returning lowest-cost nodes in an efficient way.
(iii) The function get_single_veh_availability() maps a vehicle position to a node if it is on one or to the next node if it is on an edge. In the latter case, the available time is extrapolated from the current time, the relative position on the current edge and the travel time of the current edge. Hence, the Dijkstra routing algorithm does not have to consider all possible positions on the edges.
(iv) The output of the function get_single_veh_availability() depends on the *locked* status of plan stops. Without any locked plan stops, the current location and the current time determine the node and time of availability. For a hailing operation, it makes sense to lock a drop-off stop when a customer boards a vehicle as the drop-off stop has to follow the pick-up. In this case, the availability is set to the drop-off node and the extrapolated arrival time.
(v) The general vehicle availability does not have to be updated for every single request. It is sufficient to perform it once per time step. If locked assignments were added to a vehicle, it would suffice to update the availability of this single vehicle.

Algorithm 2: search_available_vehicles

Data: G; V; t; r;
Result: V_r = search_available_vehicles$(V, r; t; G)$

```
/* general vehicle availability /*
d_av ← {} ;                                   /* mapping of search nodes to list of vehicles */
for v ∈ V do
    (n_av, t_av) ← get_single_veh_availability(v);
    (t_av, v) → d_av[n_av] ;                  /* add vehicle information to node */
end
N_v ← d_av.keys() ;                           /* destinations of Dijkstra search */

/* request-specific route search /*
V_r ← ∅ ;
c^max ← τ_r^w − t;
f ← PriorityQueue() ;                         /* PQ as frontier data structure */
N_e ← ∅ ;                                     /* set of explored nodes */
next ← {} ;                                   /* mapping of node to next node on routes */
cost ← {} ;                                   /* mapping of node to total costs from source */
(x_r^p, 0) → f ;                              /* add source node with zero cost to frontier */
cost[x_r^p] ← 0;
while f is not empty do
    n_act, c_act ← f.remove_min_element_and_costs();
    if c_act > c^max then
        break;
    end
    if n_act ∈ N_e then
        continue ;                            /* PQ elements are not invalidated */
    end
    if n_act ∈ N_v then
        for (t_av, v) ∈ d_av[n_act] do
            if t_av + c_act ≤ τ_r^w then
                v → V_r ;                     /* add v to set of available vehicles */
            end
        end
    end
    for (n_next, c) ∈ n_act.get_neighbors_with_costs do
        if n_next ∉ N_e then
            c_new ← c_act + c;
            c_old ← cost.get(n_next, ∞) ;     /* infinite if not found */
            if c_new < c_old then
                (n_next, c_new) → f ;         /* PQ update */
                cost[n_next] ← c_new;
                next[n_next] ← n_act;
            end
        end
    end
    n_act → N_e
end
```

One-to-One Routes for Existing Travel-Time Matrix

Let $G = (N, E)$ be a network graph and $C : E \to \mathbb{R}$ be the costs on the edges. Moreover, let M be the travel cost matrix, where M_{ij} represents the costs of traveling from node $i \in N$ to node $j \in N$ along the least-cost path in G. This matrix can be pre-processed by various algorithms, for instance the Floyd-Warshall algorithm.

The memory to save the matrix M scales with $\mathcal{O}(N^2)$. Hence, saving M is feasible for networks with a few thousand nodes. The memory to save the actual least-cost routes scales with $\mathcal{O}(N^3)$ (as the route length scales with N) and becomes a problem for much smaller networks. The following algorithm describes a method to re-create the least-cost path ϕ between two arbitrary nodes $o \in N$ and $d \in N$ with the help of Matrix M, which is by far more efficient than Dijkstra. It can be viewed as an A^* with perfect guidance.

Algorithm 3: get_route_with_matrix

```
Data: o, d ∈ N; G = (N, E); M;
Result: ψ = get_route_with_matrix(o, d)
φ ← [o] ;                                          /* route (list of nodes) */
c ← o ;                                            /* current node */
while c! = d do
    l_c ← get_next_nodes(c) ;                      /* list of neighboring nodes */
    for n ∈ l_C do
        if M_cd = M_cn + M_nd then
            φ ← φ + [n] ;                          /* add next node to route */
            c ← n;
            break for loop
        end
    end
end
```

In a practical implementation with float matrix values, $M_{cd} = M_{cn} + M_{nd}$ should be replaced with $M_{cd} - (M_{cn} + M_{nd}) \leq \epsilon$ with a small $\epsilon > 0$ to account for rounding effects.

B.3 Pavone's Repositioning Algorithm

[M. Pavone et al., 2012] introduced a real-time rebalancing policy, which serves as baseline in this thesis. The number β_{od} of vehicles to be rebalanced between two zones o and d is determined from an ILP.

$$\min_{\beta_{od}} \sum_{od \in Z^2} t_{od} \cdot \beta_{od} \tag{1a}$$

$$\text{s.t.} \sum_{z' \neq z} (\beta_{z'z} - \beta_{zz'}) \geq v_z^d - v_z^e \qquad \forall z \in Z \tag{1b}$$

$$\beta_{od} \geq 0 \qquad \forall od \in Z^2 \tag{1c}$$

Here, t_{od} is the expected travel time between zones o and d. In this problem, v_z^e denotes the number of excess vehicles in zone z. In this thesis, the determination of this value is modified from the original paper to consider forecasts of arriving ($\bar{\Lambda}_z$) and departing (Λ_z) vehicles in a single time horizon. The excess vehicles are computed for each zone $z \in Z$ by

$$v_z^e = \max\left(0, |V_z^I| + \min\left(0, \bar{\Lambda}_z - \Lambda_z\right)\right) \qquad \forall z \in Z \tag{2}$$

Since repositioning trips should start at the time step, in which this problem is solved, the number of repositioning vehicles departing from a zone should be smaller than $|V_z^I|$, the number of idle vehicles in zone z. Hence, the number of excess vehicles, which represent the maximum amount of vehicles that can be repositioned from a zone is bound by $[0, |V_z^I|]$.

The approach by [M. PAVONE et al., 2012] is to balance the excess vehicles in all zones, i.e.

$$v_z^d = \left\lfloor \frac{\sum_{z \in Z} v_z^e}{|Z|} \right\rfloor \qquad \forall z \in Z \tag{3}$$

B.4 Demand-Responsive Zone Creation Algorithm

Let $G = (N, E)$ be a network, where each node $n \in N$ has coordinates (x_n, y_n) in a metric system. Moreover, let T^{OD} be a set of (expected) OD trips (with $OD \in N \times N$). The following algorithm creates quadratic zones, where the edge length of the zones depends on the amount of (expected) demand within the respective zones. The algorithm requires the maximum edge length l_{max} and the number of zoom levels n_l as input. It returns a mapping Z of zone-indices to geometrical shapes and a list of zone-indices L that make up the final zone system. The procedure is divided into two steps. First, all zones for all zoom levels are created in Algorithm 4. Then, the demand-based selection of zones is conducted recursively in Algorithm 5.

Algorithm 4: create_hierarchical_zones

Data: $l_{max} > 0; n_l > 0; G = (N, E); T^{OD}$;
Result: Z = create_hierarchical_zones(l_{max}, n_l, G, T^{OD})
$x_{min}, x_{max}, y_{min}, y_{max} \leftarrow$ get_boundaries(G, T^{OD}, l_{max});
$T_{tot} \leftarrow$ get_total_demand(T^{OD});
$S = \{\}$; /* mapping of zone-index to share of total demand */
$Z = \{\}$;
$l \leftarrow 0$; /* resolution/zoom level */
while $l < n_l$ **do**
  $l_l \leftarrow l_{max}/2^l$; /* edge length of zoom level l */
  $n_x^l \leftarrow (x_{max} - x_{min})/l_l$;
  $n_y^l \leftarrow (y_{max} - y_{min})/ \cdot l_l$;
  $n_x \leftarrow 0$; /* x-index of zone on current level */
  while $n_x < n_x^l$ **do**
    $n_y \leftarrow 0$; /* y-index of zone on current level */
    while $n_y < n_y^l$ **do**
      $i \leftarrow (l, n_x, n_y)$; /* zone index */
      $z \leftarrow$ create_zone_area(l_l, n_x, n_y);
      $d \leftarrow$ get_demand_in_area(z, G, T^{OD})$/T_{tot}$;
      if $d > 0$ **then**
        $Z[i] \leftarrow z$;
        $S[i] \leftarrow$ d;
      end
      $n_y \leftarrow n_y + 1$;
    end
    $n_x \leftarrow n_x + 1$;
  end
  $l \leftarrow l + 1$;
end

Here, get_boundaries(G, T^{OD}) returns the geometrical boundaries of nodes with demand. The boundary values are selected such that $x_{max} - x_{min}$ and $y_{max} - y_{min}$ are dividable by

l_{max} and the additional length to reach this integrality condition is added symmetrically to the lower and upper end. The function create_zone_area(l_l, n_x, n_y) builds a square with edge length l_l around the center point $(x_{min} + l_l * (n_x + 0.5), y_{min} + l_l * (n_y + 0.5))$. The function get_demand_in_area(z, G, T^{OD}) counts the (expected) number of trip departures and arrivals within zone z.

For the second algorithm, an additional input parameter is required to determine how large the share of demand s_{min} should be within a zone to replace this zone with the zones of the next zoom level (if these are available). The recursive algorithm is initiated with the zones of the lowest resolution, i.e. $Z_c = \{i \in Z : i[0] = 0\}$.

Algorithm 5: select_hierarchical_zones

```
Data: Z_c; n_l; S; s_min > 0;
Result: L = select_hierarchical_zones(Z_c, n_l, S, s_min)
L ← [];
for i ∈ Z_c do
    (l, n_x, n_y) ← i;
    if S[i] < s_min OR l == n_l then
        ; /* add current zone index */
        L ← L + [i]
    else
        ; /* check zones of next zoom level */
        Z_c^p ← {};
        I_nl ← [(l + 1, 2n_x, 2n_y), (l + 1, 2n_x + 1, 2n_y), (l + 1, 2n_x, 2n_y + 1), (l + 1, 2n_x + 1, 2n_y + 1)];
        for i_nl ∈ I_nl do
            Z_c^p[i] ← Z[i];
        end
        L ← L+ select_hierarchical_zones(Z_c^p, n_l, S, s_min)
    end
end
```

For the purpose of creating zones with sizes based on demand, Algorithms 4 and 5 can be viewed as an alternative to machine-based clustering methods (e.g., K-Means). In contrast to K-Means, this approach creates regular geometric shapes and considers constraints for maximum and minimum edge length.

B.5 FleetPy

When simulating mobility-on-demand fleets, modeling detail is of essence. Simpler models typically have a higher computational performance, but might lack the capability to answer specific research questions. After adapting and restarting several frameworks from scratch, FleetPy was designed and developed by the author of this thesis and Roman Engelhardt, with the intention to build a modular framework with clear interfaces for different agents and interactions in the real and digital world. The framework proved useful for other PhD students leading to more contributors. Current development includes different aspects of the framework, e.g., dynamic access points for pooling, non-myopic user-assignment, repositioning, charging, and the synchronization of parcel and people delivery.

A first version with the base code was published as open-source code under the MIT license and is available at https://github.com/TUM-VT/FleetPy. As mentioned, the FleetPy project is still ongoing and under development.

C Inputs for Case Study in the AMoD Regulation Chapter

Chapter	Symbol	Description	Unit	Case Study Value	Source
Bayesian	$\mathcal{H}_\zeta$	Bessel function			
Optimization	ζ	Order of Bessel function		$5/2$	
Traveler	c^{VOT}	value of time (same for each mode)	€ /s	-0.0045	from[1]
Model	$c^{D,PV}$	distance dependent private vehicle cost	€ /km	0.66	from[2]
	c^{PV}	private vehicle mode choice intercept for modal calibration	€	4.70	
	f_{od}^{PT}	public transport fare for traveling from origin o to destination d	€	1.00	
	v^{walk}	walking speed	m/s	1.33	
	c^T	public transport transfer penalty	€	-0.8	
	$(\eta_1, g(\eta_1))$	first point of definition of $g^{VOT}(\eta)$		(0,1)	
	$(\eta_2, g(\eta_2))$	second point of definition of $g^{VOT}(\eta)$		(0.38,1)	from[3]
	$(\eta_3, g(\eta_3))$	third point of definition of $g^{VOT}(\eta)$		(1,1.76)	from[4]
	$(\eta_2, g(\eta_2))$	forth point of definition of $g^{VOT}(\eta)$		(1,10000)	
Street	$v_{1,I}$	first travel time scaling parameter for inner zone	m/s	5.87	
Network	$v_{2,I}$	second travel time scaling parameter for inner zone	m/s	7.37	
Model	$v_{1,O}$	first travel time scaling parameter for outer zone	m/s	10.32	
	$v_{2,O}$	second travel time scaling parameter for outer zone	m/s	15.35	
Public	ρ^{Bus}	assumed capacity for bus	pax	100	from[3]
Transport	ρ^{Tram}	assumed capacity for tram	pax	216	from[3]
Network	ρ^{Subway}	assumed capacity for subway train	pax	940	from[3]
Model	$\rho^{UrbanTrain}$	assumed capacity for urban train	pax	1088	from[3]
	c^{Bus}	cost for bus	€ /km	3.00	from[5]
	c^{Tram}	cost for tram	€ /km	5.00	from[5]
	c^{Subway}	cost for subway train	€ /km	15.00	from[5]
	$c^{UrbanTrain}$	cost for urban train	€ /km	15.00	from[5]
	w^{Bus}	energy consumption for bus	kWh/km	34	from[6]
	w^{Tram}	energy consumption for tram	kWh/km	26	from[6]
	w^{Subway}	energy consumption for subway	kWh/km	113	from[6]
	$w^{UrbanTrain}$	energy consumption for urban train	kWh/km	173	from[6]
	e_w	emission of CO2 per energy consumption	g/kWh	112	from[7]
Fleet	τ_r^w	customer waiting time constraint	s	30	
Control	τ_r^δ	detour time constraint			
Model		relative to direct route travel time	%	40	
	t^b	boarding duration	s	30	
	ρ	AMoD vehicle capacity	pax	4	
	$(N^{RV,wl}, N^{RV,al})$	number of vehicles considered by RV heuristic	veh	(5,5)	
Modeling	x^P	regulator decision variable for parking fees	€	[2.50, 5.00]	

[1] [Frei et al., 2017]
[2] https://www.tcs.ch/de/testberichte-ratgeber/ratgeber/kontrollen-unterhalt/kilometerkosten.php
[3] https://www.mvg.de/ueber/das-unternehmen/fahrzeuge.html
[4] [Tirachini, Hurtubia, et al., 2017]
[5] https://prof.beuth-hochschule.de/fileadmin/prof/jschlaich/200811_Fr_JS_Kostenmodelle_-NAHVERKEHR.pdf
[6] https://www.ris-muenchen.de/RII/RII/DOK/ANTRAG/2337762.pdf
[7] https://www.swm.de/dam/swm/dokumente/geschaeftskunden/broschuere-strom-erdgas-gk.pdf

Chapter	Symbol	Description	Unit	Case Study Value	Source
the Impacts	x^{RT}	regulator decision variable for road toll	€ /km	[0.0, 1.0]	
of Decision	k_0	threshold macroscopic density to trigger road toll	veh /lane-km	5.0	
Variables	x^{PT}	regulator decision variable for scaling public transport frequency		[0.25, 2.0]	
	x^{F}	regulator decision variable to limit the AMoD fleet size	vehicles	[0, 50,000]	
	y^{F}	AMoD service planner decision variable for fleet size	vehicles	[0,10000]	
	y^{PD}	AMoD service planner decision variable for distance-based fare	€ /km	[0.25, 2.00]	
	f_{min}^{AMoD}	minimum base fare for an AMoD trip	€	1.0	
	y^{PU}	AMoD service planner decision variable to scale fares if 75% fleet utilization is exceeded		[1.0, 10.0]	
Social	e^{PV}	avg private vehicle CO2 emissions	g/km	130	from[8] [9]
Welfare	e^{AMOD}	avg AMoD vehicle CO2 emissions	g/km	15	from[10][11]
and Profit	c^{CO2}	monetarization of CO2 emissions	€ /kg	0.145	from[12]
Model	c_v^F	(daily) fixed costs of an AMoD vehicle	€	40	from[13]
	c_v^D	distance dependent operating cost of an AMoD vehicle	€ /km	0.25	from[13]

Table A2: Collection of parameters within the case study in the AMoD regulation chapter sorted by their first occurrence. Fixed parameter values within the case study are given with their source.

D Overview-Tables for Literature Review

For the sake of legibility, abbreviations are used for most categories, which are explained in the respective paragraphs in Chapter 2.

Abbreviations in Table A3:

- **Operating Area**: B: benchmark instance, G: grid network instance, O: optimization of operating area, R: random instance
- **Charging Infrastructure**: GP: generation procedure, N: variation in number of chargers, O(L): optimization (layer), S: battery swap
- H: hailing, P: pooling
- **Fleet Composition**: B: battery-size variation, H: homogeneous, M: mixed fleet

[8] https://www.muenchen.de/rathaus/Stadtinfos/Statistik/Verkehr.html
[9] https://www.ris-muenchen.de/RII/RII/DOK/ANTRAG/2337762.pdf
[10] https://ev-database.uk
[11] https://www.swm.de/dam/swm/dokumente/geschaeftskunden/broschuere-strom-erdgas-gk.pdf
[12] http://www.suedumfahrung-jetzt.de/wp-content/uploads/2016/04/bvwp-2030-methodenhandbuch.pdf
[13] https://www.tcs.ch/de/testberichte-ratgeber/ratgeber/kotrollen-unterhalt/kilometerkosten.php

- **Pricing**: A: auction, D: discount for pooled trips, F: fix price structure, O: output of cost analysis, PO: price optimization, S: spatial pricing, T: temporal pricing, VEC: variable electricity costs
- **Customer-Operator Interaction**: AS: asynchronous information flow, B: booking, C-A/R: customer acceptance/rejection, O-A/R: operator acceptance/rejection, R: request, RT: retry
- **Time Constraints**: CD: Common Destinations, D: Detour Time, DTU: Derived from Travel Utility, R: Reservation Time, SD: Schedule Deviation, TTD: Total Travel Delay, PTW: Pickup Time Window, W: Wait Time
- **MaaS/Intermodal Integration**: F: Feeder, I: Indirect (Analysis of Lacking PT Connections), JP: Joint Planning, N: Network

Abbreviations in Table A4:

- **Network**: G: grid network, E: Euclidean plane, M: Manhattan plane, MS: main street network, S: street network, VH: virtual hub network, VHC: virtual hub cascade
- **Customer & Vehicle Model**: F: flow model, Q: quantization of flows, A: agent-based model, M: macroscopic model
- **Travel Times & Routing**: -: routing not considered/necessary, BPR/BPM: routing based on link-flow dependent travel times (BPR/breakpoint model), CS: routing with constant speed, CTM: routing with cell transmission model, E: routing based on Euclidean distance, FCM: routing with flow-capacity model, IM: routing based on travel time matrices from iterative simulations, M: routing based on travel time matrix, RS: routing with random speed, S: routing based on and movement in traffic simulation, TM: routing based on time-dependent travel time matrices
- **PT Cooperation Model**: xNS: consideration of hyperpaths through x nearest PT stations, LCH: preprocessing of least-cost hyperpaths, WNC: consideration of walking network connectivity, NSS: feeder service only near selected stations
- **AMoD Competition**: x Op: competition of x operators, H/Px: mixed fleet with different options (hailing, pooling with max. occupancy of x)
- **AMoD Demand**: AB: activity-based model, DL: day-to-day learning of agents, E: exogenous demand, HB: hierarchical Bayesian model, MC: mode choice model, MC-AS: asynchronous mode choice model, PS: price sensitive demand, WS: waiting-time sensitive demand
- **Stochasticity in Customer Interaction**: D: Deterministic, P: Probabilistic

Abbreviations in Table A5:

- **Fleet Sizing**: D: dynamic vehicle creation in simulation, EDD: elastic data driven fleet sizing, M: macroscopic estimation, O: optimization

- **Customer-Vehicle Assignment**: cDSO: combined Dynamic State Optimization, cMPC: combined Model Predictive Control, DSO: Dynamic State Optimization, DSO-AS: Asynchronous Dynamic State Optimization, DSO-ML: Machine Learning for DSO, DTI: Dwell-Time Insertion, IH: Insertion Heuristic, S: Shareability Estimation, SO: Static Optimization, SQM: Station Queue Model, ST: Schedule Transition, VSH: Vehicle Selection Heuristics
- **Dynamic Pricing**: cMPC: combined Model Predictive Control
- **Repositioning**: BB: block-balance algorithm, cDSO: combined dynamic state optimization, cMPC: combined model predictive control, DH: dual horizon approach, DSO: dynamic state optimization, EMR: early-morning repositioning, EQ: using equilibrium model, FB: feedback control, FF: feed-forward control, HR: using hypothetical requests, LO: applying local optimization such as search of next depot, MPC: model predictive control approach, ODA: only demand anticipation, OSA: only supply anticipation, RC: random cruising, RFM: using realistic non-perfect demand forecast methods, RL: using reinforcement learning methods, RR: using rejected requests as demand estimation
- **Charging**: cMPC: combined model predictive control, EP: electricity price consideration, ID: charging while idle at depot, MP: macroscopic planning model, MPC: model predictive control, PN: power network considerations, ST: static schedule transition, T: threshold criterion, TR: trip rejection criterion

Citation	Case Study Operating Area(s)	Charging Infrastructure	Hail (H) / Pool (P)	Fleet Composition	Pricing considered	Customer Interaction	Time Constraints	MaaS & PT Integration
Al-Kanj et al. [2020]	New Jersey	100 kW	H	H	**S,T**	**R,O-A/R,C-A/R**	-	-
Atasoy et al. [2015]	Hino (Tokio)	-	H,P	**M**	F	**R,Broker,C-A/R**	SD	**Broker**
Basu et al. [2018]	Virtual City (Singapore)	-	P	H	F	B	W,D	**F**
Bauer et al. [2018]	Manhattan	**N,P**	H	H,**B**	O	B	W	-
Bilali, Dandl, et al. [2019b] and Bilali, Dandl, et al. [2019a]	Munich	-	P	H	-	B	**R,W,D,B**	-
Bilali, Engelhardt, et al. [2020]	Munich	-	P	H	-	B	**R,W,D,B**	-
Bischoff, Kaddoura, et al. [2018]	**O**, Berlin	-	P	H	-	R,O-A/R	W,D	-
Bongiovanni et al. [2019]	B	3 kW	P	**M**	-	R,O-A/R	TW	-
T. D. Chen and Kockelman [2016]	G	**GP**,4/50 kW	H	H,**B**	O	R,C-A/R	W	-
Dandl and Bogenberger [2018]	Munich	-	H	**M**	-	R,O-A/R	W	-
Dandl, Bogenberger, and Mahmassani [2019]	Manhattan	-	H	H	F	**AS,R,O-A/R,C-A/R**	-	-
Dandl, Grueber, et al. [2019]	Munich	-	P	H	O	**B,RT**	**R,W,D**	**I**
Dandl, Fehn, et al. [2020]	Munich	**N**,163 kW	H	H	**VEC**,O	B	-	-
Engelhardt, Dandl, Bilali, et al. [2019] and Engelhardt, Dandl, and Bogenberger [2019]	Munich	-	P	H	-	R,O-A/R,C-A/R	PTW,D	-
Erdmann, Dandl, Kaltenhaeuser, et al. [2020]	Manhattan	-	H	H	-	R,O-A/R	**W,R**	-
Sebastian Hörl et al. [2019]	Paris	-	H	H	**PO**	B	-	-
Iacobucci et al. [2019]	Tokyo	20 kW	H	H	**VEC**	B	-	-
Jung, J. Y. Chow, et al. [2014]	Seoul	**OL**	P	H	-	R,O-A/R	W,D	-
Kang et al. [2015]	Ann Arbor	**OL**	H	H	OL	B	-	-
Kucharski and Cats [2020]	Amsterdam	-	P	H	**D**	-	U	-
Liang, G. H. d. A. Correia, et al. [2016]	**O**,Delft	1.5 kW	H	H	O	R,O-A/R	-	I
Y. Liu, Bansal, et al. [2019]	Manhattan	-	H,P	**M**	**D**	R,O-A/R	W,TTD	-
Ma et al. [2019]	Long Island	-	P	H	-	B	-	**F**
Martinez et al. [2015]	Lisbon	-	P	H	-	B	**R,W,TTD**	-
Nahmias-Biran et al. [2020]	Virtual City (Singapore)	-	H,P	H	F,**D**	B	-	**F**
Oke, Akkinepally, et al. [2020]	Baltimore, Boston	-	H,P	H	F,**D**	B	-	**F**
Pinto et al. [2019]	Chicago	-	P	H	F	B	D	**JP**
Salazar, Rossi, et al. [2018]	Manhattan	-	H	H	**S**	B	-	**N**
Shen et al. [2018]	(part of) Singapore	-	H,P	H	F,**D**	R,O-A/R	W,D	**F**
Sheppard et al. [2019]	**9 study areas in US**	10-250 kW	H	H,**B**	O	B	W	-
Sieber et al. [2020]	4 rural areas Switzerland	-	H	H	O	B	-	**F**
Tachet et al. [2017]	**11 Cities**	-	P	H	-	B	TTD	-
Turan et al. [2020]	Manhattan, San Francisco	48 kW	H	H	**S,T,VEC**	B	-	-
Vakayil et al. [2017]	Washington D.C.	-	H	H	-	R,O-A/R	W	**F**
Vosooghi, Puchinger, Jankovic, et al. [2019]	Rouen Metro Area (France)	-	P	H	F,**D**	R,O-A/R	W,D	-
Vosooghi, Puchinger, Bischoff, et al. [2020]	Rouen Metro Area (France)	O,22/43 kW,S	P	H,**B**	F	B	-	-
Wallar, Schwarting, et al. [2019]	Manhattan, Singapore	-	P	**M**	-	B	W,TTD	-
Wen, Y. X. Chen, et al. [2018]	Suburban Area in Europe	-	P	H	F,**D**	R,O-A/R	**R,W,D**	F
Wilkes et al. [2021]	Small town in Germany	-	P	H	F	**R,O-A/R,C-A/R**	W,D	-
J. J. Q. Yu et al. [2018]	R	-	H,P	H	**A**	**R,Broker**	-	**Broker**
Zachariah et al. [2014]	New Jersey State	-	P	H	-	B	W,D,**CD**	**F**
Zgraggen et al. [2019]	Manhattan	-	H	H	-	R,O-A/R	-	**N**
H. Zhang et al. [2019]	B, Long-Distance	**O**,100 kW	H	H	F	B	-	-
Wenwen Zhang et al. [2015]	R	-	P	H	**D**	B	-	-
Zhu et al. [2017]	New Jersey	-	P	**M**	-	B	-	-

Table A3: Classification of papers regarding AMoD service design. Papers are mentioned if part of their focus is on service design or if their service design differs in at least one category from most of the other papers. This category is marked (bold letters).

Citation	Study Focus	Network Model	Customer & Vehicle Model	Travel Times & Routing	AMoD Competition	PT Cooperation	AMoD Demand	Stochasticity of Customer Interaction
Al-Kanj et al. [2020]	O	G,**VHC**	A	-	-	-	**PS**	**P**
Atasoy et al. [2015]	O,D	S	A	M	H,P8	-	**MC**	**P**
Azevedo et al. [2016]	O,D	VH, S	A	**IM**	-	-	**AB,MC**	D
Basu et al. [2018]	O,D	MS	A	IM	-	**NSS**	**DL,AB,MC**	D
Bilali, Dandl, et al. [2019b] and Bilali, Dandl, et al. [2019a]	O	E	**M**	-	-	-	E	D
Bilali, Engelhardt, et al. [2020]	O	E,S	**M,A**	M	-	-	E	D
Burns et al. [2013]	O	E	**M,A**	-	-	-	E	D
Daganzo and Ouyang [2019]	O	E	**M,A**	-	-	-	E	D
Dandl, Bracher, et al. [2017]	O,**T**	S	A	**S**	-	-	E	D
Dandl, Bogenberger, and Mahmassani [2019]	O	M	A	CS	**2 Op**	-	**MC-AS**	**P**
Dandl, Fehn, et al. [2020]	O	E	**M**	-	-	-	E	D
Engelhardt, Dandl, Bilali, et al. [2019]	O	S	A	TM	-	-	**WS**	**P**
Fagnant and Kockelman [2014]	O	M,**VHC**	A	CS	-	-	E	D
Fiedler et al. [2018]	O	S	A	**E,CS**	-	-	E	D
Fluri et al. [2019]	O	S,**VHC**	A	S	-	-	E	D
Herminghaus [2019]	O	E	**M**	CS	-	-	PS	D
S. Hörl et al. [2019]	O	S	A	**S, E**	-	-	E	D
Kang et al. [2015]	O	E	A	CS	-	-	**HB**	D
Levin et al. [2017]	O,**T**	S	A	**CTM**	-	-	E	D
Levin [2017]	O,**T**	S	F	**CTM**	-	-	E	D
Q. Li and Liao [2020]	O,D	S,VH	F	**BPR**	-	-	AB,DUE,SO	D
Liang, G. H. d. A. Correia, et al. [2018]	O	S	A	**BPM**	-	-	E	D
Liang, Correia, Gonçalo Homem de Almeida, et al. [2020]	O	S	A	**BPR**	-	-	E	D
Loeb and Kockelman [2019]	O	S	A	IS	-	-	**WS**	**P**
Ma et al. [2019]	O	E	A	CS	-	**4NS**	E	D
Maciejewski and Bischoff [2017]	O,**T**	S	A	**S**	-	-	E	D
Marczuk et al. [2015]	O,D,**T**	S	A	**S**	-	-	AB,MC	D
Oke, Akkinepally, et al. [2020]	O,D	S	A	IM	-	**7.5mi**	DL,AB,MC	D
Pinto et al. [2019]	O,D	M?	A	CS	-	**LCH**	MC	D
Ruch, Horl, et al. [2018]	O	S,VH	A	**S**	-	-	E	D
Salazar, Rossi, et al. [2018]	O	S	F	FCM	-	**WNC**	E	D
Salazar, Tsao, et al. [2019]	O	S	F	**BPR**	-	-	E	D
Séjournè et al. [2018]	O	VH	F	M	**2 Op**	-	E	D
Simonetto et al. [2019]	O	S	A	M	**2 Op**	-	E	D
Turan et al. [2020]	O	VH	A	M	-	-	**PS**	**P**
Vakayil et al. [2017]	O	VH	A	M	-	**LCH**	E	D
Vazifeh et al. [2018]	O	S	A	TM	**1-3 Op**	-	E	D
Wilkes et al. [2021]	O,D	S	A	M	-	-	**AB,MC**	**P**
Z. Xu et al. [2019]	O	E	**M**	CS	-	-	E	D
J. J. Q. Yu et al. [2018]	O	-	A	R	**Multiple**	-	**PS**	**P**
Zgraggen et al. [2019]	O	VH,PT	F & Q	M	-	**WNC**	E	D

Table A4: Classification of papers regarding AMoD modeling in transportation systems. Papers are mentioned if part of their focus is on the development of AMoD modeling. This category is marked (bold letters).

Citation	Fleet Sizing	Customer Assignment	Dynamic Pricing	Repositioning	Charging
Albert et al. [2019]	400-1000	SQM	-	**MPC**	-
Al-Kanj et al. [2020]	1000-2200	**cMPC**	**cMPC**	**cMPC**	**cMPC**
Alonso-Mora, Samaranayake, et al. [2017]	1000-3000	**DSO**	-	DSO,**RR**	-
Alonso-Mora, Wallar, et al. [2017]	1000-3000	cDSO,**HR**	-	cDSO,**HR**	-
Atasoy et al. [2015]	**O**, 60	IH	-	-	-
Bauer et al. [2018]	D	ST	-	-	**SST**
Bertsimas et al. [2019]	2000-10000	**IH, DSO**	-	-	-
Bilali, Dandl, et al. [2019b] and Bilali, Dandl, et al. [2019a]	-	**S**	-	-	-
Bilali, Engelhardt, et al. [2020]	100-1500	**S, DSO**	-	-	-
Bischoff and Maciejewski [2016b] and Bischoff and Maciejewski [2016a]	**50000-250000**	IH	-	-	-
Boesch et al. [2016]	4000-130000	**ST**	-	-	-
Boewing et al. [2020]	10-45	**cMPC**	-	-	**cMPC,PN**
Bongiovanni et al. [2019]	16-50	**SO**	-	-	**SST**
Carron et al. [2019]	4000	SQM	-	**MPC**	-
Charkhgard et al. [2020]	5-100	SQM	-	**MPC**	-
T. D. Chen and Kockelman [2016]	30000-60000	IH	-	BB	**TR**
Daganzo and Ouyang [2019]	1-12000	**M,IH**	-	-	-
Dandl and Bogenberger [2018]	50-4000	**DSO, IH**	-	-	-
Dandl and Bogenberger [2019]	50-300	DSO	-	**DSO**	T
Dandl, M. Hyland, et al. [2019]	5000	**cDSO**	-	**cDSO,RFM**	-
Dandl, Bogenberger, and Mahmassani [2019]	3500-4500	**AS-DSO**, IH	-	-	-
Dandl, Grueber, et al. [2019]	1-100	**ST**	-	-	-
Dandl, M. Hyland, et al. [2020]	M,2750-4500	**cDSO**	-	**cDSO,DH**	-
Dandl, Fehn, et al. [2020]	2000-2500	**M**	-	-	**MP,MPC,EP**
Engelhardt, Dandl, Bilali, et al. [2019] and Engelhardt, Dandl, and Bogenberger [2019]	200-3000	**VSH,DSO**	-	-	-
Erdmann, Dandl, Kaltenhaeuser, et al. [2020]	50-300	**R-IH**	-	-	-
Estandia et al. [2021]	16000	SQM	-	**cMPC**	**cMPC,PN**
Fagnant and Kockelman [2014]	**D**,400-3300	IH	-	**BB**	-
Fagnant and Kockelman [2018]	**O**,1500-2500	**IH**	-	BB	-
Farhan, D. Chen, et al. [2018]	45-80	**IH**	-	-	**ID**
Farhan and T. D. Chen [2018]	15000-55000	**IH**	-	BB	TR?
Fehn et al. [2019]	5-30	IH	-	-	**T,EP**
Fluri et al. [2019]	-	DSO	-	**RL**	-
Herminghaus [2019]	**M**	**M**	-	-	-
Hu and J. Dong [2020]	650	DSO, **ML-DSO**	-	-	T
M. Hyland and Mahmassani [2018a]	100-800	IH, **DSO**	-	-	-
M. Hyland, Dandl, et al. [2019]	5000	**cDSO**	-	**cDSO,RFM**	-
M. F. Hyland and Mahmassani [2020]	400	**DSO**	-	-	-
Iacobucci et al. [2019]	30	MPC	-	MPC	**MPC**
Jaeger et al. [2018]	45000	**IH**	-	-	-
Jung, J. Y. Chow, et al. [2014]	600	**IH**	-	-	T
Kondor et al. [2020]	**D**,1-400000	IH,S	-	-	-
Kucharski and Cats [2020]	-	**SO**	-	-	-
L. Li, Lin, et al. [2019]	200-2000	IH	-	NaN	**ID, T**
L. Li, Pantelidis, et al. [2021]	262,1400-2000	IH	-	-	**TR**
Q. Li and Liao [2020]	O,2000-10000	-	-	**EQ**	-
Liang, G. H. d. A. Correia, et al. [2018]	5-25	**DSO**	-	-	-
Liang, Correia, Gonçalo Homem de Almeida, et al. [2020]	5-25	**DSO**	-	-	-
Y. Liu, Bansal, et al. [2019]	O,1000-4000	**DSO**	-	RR	-
Y. Liu and Samaranayake [2019]	1000-3000	DSO	-	**HR**	-

Citation	Fleet Sizing	Customer Assignment	Dynamic Pricing	Repositioning	Charging
Loeb, Kockelman, and J. Liu [2018]	12000-39000	ST	-	-	**T,TR,ID**
Ma et al. [2019]	400-2200	**IH**	-	**DSO**	-
Maciejewski, Bischoff, and Nagel [2016]	500-2500	**IH**	-	-	-
Navjyoth Sarma et al. [2020]	500-3000	**DSO,IH**	-	-	-
M. Pavone et al. [2012]	50-150	SQM	-	**FB,FF**	-
Pinto et al. [2019]	**O**, 250-1500	IH	-	-	-
Ruch, Lu, et al. [2020]	15-600	**IH, DSO**	-	RR,BB	-
Santi et al. [2014]	-	**S**	-	-	-
Simonetto et al. [2019]	150-3000	**VSH,DSO**	-	RR	-
Syed, Akhnoukh, et al. [2019]	10	**IH**	-	-	-
Syed, Kaltenhaeuser, et al. [2019]	50-200	**IH**	-	-	-
Tachet et al. [2017]	-	**S**	-	-	-
Turan et al. [2020]	400-1200	SQM	**cMPC**	**cMPC**	**cMPC,ID,EP**
Vazifeh et al. [2018]	6000-12000	**DSO,ST**	-	-	-
Vosooghi, Puchinger, Bischoff, et al. [2020]	3000	IH	-	-	**T,TR**
Wallar, van der Zee, et al. [2018]	1000-3000	DSO	-	**DSO**	-
Wallar, Schwarting, et al. [2019]	O	**ST**	-	-	-
S. Wang et al. [2019]	2000-4500	**IH,DSO**	-	-	-
Wen, Zhao, et al. [2017]	20-300	IH	-	BB,**DSO,RL**	-
Wilkes et al. [2021]	25-150	**IH, DSO**	-	-	-
Winter, Cats, G. Correia, et al. [2018]	250-900	**DTI**	-	-	-
Winter, Cats, Martens, et al. [2020]	12500	IH	-	**LO,ODA,OSA**	-
Winter, Cats, Martens, et al. [2021]	10000-15000	IH	-	**ODA,LO**	-
Wittmann et al. [2020]	**EDD**	IH, DSO	-	**MPC**	-
Z. Xu et al. [2019]	-	**M**	-	RC	-
J. J. Q. Yu et al. [2018]	10-1000	**DSO**	-	-	-
Zgraggen et al. [2019]	4500-5500	-	-	**MPC**	-
Rick Zhang et al. [2016]	40	**cMPC**,DSO	-	**cMPC**,FB	**cMPC**
H. Zhang et al. [2019]	O	-	-	**cMPC**	**cMPC**
Zwick and Axhausen [2020]	**EDD**	IH	-	FF	-
Zwick, Kuehnel, et al. [2021]	5000-8000	**IH**	-	FF	-

Table A5: Classification of papers regarding AMoD operation categories. Papers are mentioned if part of their focus is on the development of operational aspects. This category is marked (bold letters).

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die bei der promotionsführenden Einrichtung School of Engineering and Design der TUM zur Promotionsprüfung vorgelegte Arbeit unter der Anleitung und Betreuung durch Univ.-Prof. Dr.-Ing. Klaus Bogenberger ohne sonstige Hilfe erstellt und bei der Abfassung nur die gemäß § 6 Ab. 6 und 7 Satz 2 angeboteten Hilfsmittel benutzt habe.

Ich habe keine Organisation eingeschaltet, die gegen Entgelt Betreuerinnen und Betreuer für die Anfertigung von Dissertationen sucht, oder die mir obliegenden Pflichten hinsichtlich der Prüfungsleistungen für mich ganz oder teilweise erledigt.

Ich habe die Dissertation in dieser oder ähnlicher Form in keinem anderen Prüfungsverfahren als Prüfungsleistung vorgelegt.

Die vollständige Dissertation wurde in der Schriftenreihe des Lehrstuhls für Verkehrstechnik veröffentlicht. Die promotionsführende Einrichtung Ingenieurfakultät Bau Geo Umwelt hat der Veröffentlichung zugestimmt.

Ich habe den angestrebten Doktorgrad noch nicht erworben und bin nicht in einem früheren Promotionsverfahren für den angestrebten Doktorgrad endgültig gescheitert.

Die öffentlich zugängliche Promotionsordnung der TUM ist mir bekannt, insbesondere habe ich die Bedeutung von § 28 (Nichtigkeit der Promotion) und § 29 (Entzug des Doktorgrades) zur Kenntnis genommen. Ich bin mir der Konsequenzen einer falschen Eidesstattlichen Erklärung bewusst.

Ort, Datum, Unterschrift