Technische Universität München
TUM School of Engineering and Design

Optimization of Autonomous Mobility on Demand Services based on Spatio-Temporal Relations

Arslan Ali Syed

Vorsitz: Prof. Dr.-Ing. Johannes Betz

Prüfer*innen der Dissertation: 1. Prof. Dr.-Ing. Klaus Bogenberger
2. Prof. Dr.-Ing. Jörg Ott

Abstract

The last two decades saw significant technical advancements in multiple domains, including the transport and communication sectors. The emergence of smartphones and the widespread availability of high-speed internet have revolutionized the world. This allowed the advent of widely-popular mobility on-demand (MoD) platforms where the users can request rides from origin to destination via online platforms. The MoD service then assigns a vehicle with a driver to the requested ride, and both the driver and the customer receive the option to accept or reject the ride offer. The ride could be distinguished into two types: ride-hailing (RH), where a single customer request is served at a time, or ride-sharing (RS), where multiple customers are dynamically grouped into a single ride from origins to destinations. However, encouraging people to share rides has met with limited success, and RH remains the most popular mode in MoD platforms.

With the developments in autonomous vehicle (AV) technology, the appearance of autonomous mobility on-demand (AMoD) services is also on the horizon, with some companies already in their pilot phase. The AMoD services are expected to cause another paradigm shift in the transport sector. Looking at the potential of AMoD services, there has been a drastic increase in the scientific literature to test and improve the efficiency of AMoD services even before their actual arrival into the market. A major difference between the MoD and AMoD operator is the absence of a human driver, meaning that the central fleet controller (FC) run by the AMoD operator makes all the decisions for the AMoD fleet. The two essential functions of an FC are the assignment of customer requests to AMoD fleet and the scheduling of the maintenance tasks. The latter generally consists of several types, like refueling, cleaning, and repositioning idle vehicles. Among these types, repositioning is the primary maintenance task to improve fleet utilization and AMoD profit.

Given the importance of AMoD services, the current dissertation fundamentally focuses on improving the performance of AMoD services. The present dissertation only assumes RH offers as they dominate the current MoD market. Since the FC plays the primary role in the AMoD operation, the dissertation mainly focuses on developing innovative FC methods. The dissertation provides three key contributions for this purpose, as described below.

- The **first** contribution is building a consistent relationship between the AMoD supply and demand. The supply-demand balance plays a significant role in the AMoD operation. For a high AMoD performance, the FC must ascertain a good distribution of AMoD vehicles such that AMoD customers are picked up within reasonable waiting times. Therefore, the FC needs a metric to measure the local supply-demand imbalance in different parts of the operation area. Traditionally, this is done by dividing the operation area into a disjoint set of regions. Usually, the fundamental assumption is that the regions are independent, i.e., a vehicle located in a region can only serve the customers in that specific region. However, the regions used are often small and contiguous, invalidating the above assumption. Therefore, the dissertation develops an innovative spatiotemporal relationship, referred to as supply-demand imbalance density (ID), based on the concept of reachability functions (RFs). A RF defines the outreach of a vehicle from the center point similar to the kernel functions used in kernel density estimation (KDE). KDE is often used in statistics as a non-parametric way of estimating the underlying probability density function. The ID formulation deals directly with the geographical coordinates and does

not depend on the definition of individual regions. Using ID, the offered service quality can also be plotted as a heat map.

- The **second** contribution builds on the ID function and uses it to develop three reachability function based repositioning (RFR) methods. Due to the computational challenges involved and the unavailability of individual coordinates for customer forecasts, the dissertation uses a region-based estimation of the ID in the RFR formulations. However, in contrast to the traditional methods based on the assumption of independent regions, RFR methods inherently consider the impact of repositioning decisions on the neighboring regions.

- The **third** contribution further uses the previous two contributions to improve the performance of the AMoD services. It develops a novel proactive assignment strategy that maintains the AMoD supply-demand balance while assigning AMoD vehicles to customers. This reduces the requirement for explicit repositioning of idle vehicles to demand-intensive regions and, consequently, significantly reduces the additional mileage required. The formulation uses the region-based estimation of the ID. Furthermore, it combines the novel RFR methods and the proactive assignment approach in a single AMoD operation.

To confirm the effectiveness of the above methods, the dissertation builds an agent-based simulation in the Manhattan area using the open-source New York City (NYC) taxi data. The dissertation also develops a spatiotemporal travel time scaling method to reproduce realistic network travel times from past trip data. The simulation results indicate a strong positive relationship between the ID function and the AMoD service quality offered in the operation area. Similarly, the RFR methods, the proactive assignment, and the combination of the two approaches significantly improve the AMoD performance compared to the benchmark methods used in the dissertation. This provides an essential milestone in the AMoD research and opens the door to further investigations into the topic.

List of Publications

Journal Papers

1. Arslan Ali Syed, Florian Dandl, Bernd Kaltenhäuser, and Klaus Bogenberger [June 2021]. "Density Based Distribution Model for Repositioning Strategies of Ride Hailing Services". In: *Frontiers in Future Transportation* 2, p. 681451. ISSN: 2673-5210. DOI: 10.3389/ffutr.2021.681451

2. Arslan Ali Syed, Irina Gaponova, and Klaus Bogenberger [May 2019]. "Neural Network-Based Metaheuristic Parameterization with Application to the Vehicle Matching Problem in Ride-Hailing Services". In: *Transportation Research Record: Journal of the Transportation Research Board*. ISSN: 0361-1981, 2169-4052. DOI: 10.1177/0361198119846099

3. **Planned** — Arslan Ali Syed and Klaus Bogenberger [2023]. "Imbalance Density based Proactive Assignment Method for AMoD Services". In: *Transportation Research Part C: Emerging Technologies*

Conference Papers

- Arslan Ali Syed, Florian Dandl, and Klaus Bogenberger [Sept. 2021]. "User-Assignment Strategy Considering Future Imbalance Impacts for Ride Hailing". In: *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. Indianapolis, IN, USA: IEEE, pp. 2441–2446. ISBN: 978-1-72819-142-3. DOI: 10.1109/ITSC48978.2021.9564559

- Arslan Ali Syed, Yunfei Zhang, and Klaus Bogenberger [Sept. 2023]. "Data-Driven Spatio-Temporal Scaling of Travel Times for AMoD Simulations". In: *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. Bilbao, Spain: IEEE, pp. 3583–3588. ISBN: 9798350399462. DOI: 10.1109/ITSC57777.2023.10422313

- Arslan Ali Syed, Bernd Kaltenhaeuser, Irina Gaponova, and Klaus Bogenberger [Oct. 2019]. "Asynchronous Adaptive Large Neighborhood Search Algorithm for Dynamic Matching Problem in Ride Hailing Services". In: *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. Auckland, New Zealand: IEEE, pp. 3006–3012. ISBN: 978-1-5386-7024-8. DOI: 10.1109/ITSC.2019.8916943

- Arslan Ali Syed, Karim Akhnoukh, Bernd Kaltenhaeuser, and 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, and Luís Paulo Reis. Vol. 11804. Cham: Springer International Publishing, pp. 584–595. ISBN: 978-3-030-30240-5 978-3-030-30241-2. DOI: 10.1007/978-3-030-30241-2_49

- ARSLAN ALI SYED, MARKUS FISCHER, CORNELIUS HARDT, and KLAUS BOGENBERGER [Oct. 2022]. "Charge Point Search Policies for EVs - Minimizing The Cruise for Juice". In: *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2487–2493. DOI: 10.1109/ITSC55140.2022.9921802

- YUNFEI ZHANG, ROMAN ENGELHARDT, ARSLAN-ALI SYED, FLORIAN DANDL, CORNELIUS HARDT, and KLAUS BOGENBERGER [Oct. 2022]. "Simulating Charging Processes of Mobility-On-Demand Services at Public Infrastructure: Can Operators Complement Each Other?" In: *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2200–2205. DOI: 10.1109/ITSC55140.2022.9922449

Acknowledgement

Finishing this dissertation has been an important journey for me that could only be completed with the support of multiple people. However, I begin by expressing my deepest gratitude to Allah (God), who continuously gave me all the opportunities and put these people in my life – Alhamdulillah (all praise is due to God alone). The one who does not thank people does not thank God. I would like to first thank my supervisor, Prof. Klaus Bogenberger, for allowing me to work on the topic. I highly appreciate his constant support, sharing of ideas, and valuable recommendations for progressing into the topic. Even though he became significantly busier after taking the professorship at TUM, he always made sure that his doctoral students received the necessary support required for finishing the dissertation. It has been a great pleasure to work under him and I am highly grateful to him.

Next, I would like to thank my second supervisor Prof. Jörg Ott for agreeing to supervise the thesis. I got to know him while working as an external supervisor for two master theses at his chair. Later, I was delighted that he accepted to act as my second supervisor. I am highly grateful to him for that. I would also like to thank my mentor Prof. Bernd Kaltenhäuser whose feedback on research ideas has been quite valuable. He was always available to discuss the topic and provided detailed responses to my queries.

For a major part of this dissertation, I worked as an external doctoral candidate in the mobility services department of BMW Group. I want to thank Dr. Ulrich Fastenrath who at the time worked as my manager and provided me the opportunity to work in his team. I would also like to thank my company supervisor, Irina Benkert, for sharing her ideas during my time at BMW. I also thank other colleagues in the team, especially Dr. Marvin Erdmann, who also worked on a similar topic as a doctoral candidate. Our discussions on the topic were always fruitful and he provided some crucial feedback on my writing style while preparing the dissertation draft.

I would also like to thank all my colleagues at UniBW and TUM for their valuable feedback during doctoral seminars every semester and in group meetings. I especially thank my co-authors, Dr. Florian Dandl and Roman Engelhardt; although I worked as an external doctoral candidate at the chair with comparatively less frequent in-person meetings, I learned a lot from them and deeply admire their passion and enthusiasm for the research. It was a great pleasure to work on different ideas and research papers together. I would also like to thank Dr. Simone Weikl, Yunfei Zhang, and Patrick Malcolm for their important feedback on some portions of the dissertation.

Finally, I want to thank my parents, siblings, and, especially, my wife Basma Alloush for their constant love and support throughout my time as a doctoral candidate. My mother and father played the most important role building my educational foundations and without their constant support, I would not be here working on this dissertation. During the end years of my dissertation, I could not travel enough to my parents and siblings. I am highly grateful for their support and patience in that time. Furthermore, due to the increasing trend of the home office during Covid, I was often physically present at home but mentally busy on my computer. During this phase, I could not have completed this dissertation without the support and constant encouragement of my wife. She has been the most important part of my life. In the last year of the dissertation, I was also blessed with a beautiful daughter. I am deeply grateful to my wife who mostly took care of our daughter and provided me enough time to finalize the dissertation. Even though her efforts may be invisible to masses, without her love and support this dissertation could not have been completed.

Contents

Key Terms and Abbreviations

The following table describes the main abbreviations and acronyms used in the dissertation. It also lists the page number where they are described in more detail. The following table skips the acronyms that are only used scarcely and are not necessary to understand the key contributions. The full set of abbreviations is included at the end of the dissertation.

Abbreviation	Description	Page
AMoD	Autonomous Mobility-on-Demand. A type of future mobility-on-demand service where the fleet consists of autonomous vehicles.	3
ASM	Absolute Scaling Method. A method used in the dissertation to scale the network travel times as close as possible to the recorded trip information in the data. It scales the individual network edges using linear optimization and the data of historical trips. The objective function is based on the absolute error function.	54
FC	Fleet Controller. An FC is responsible for dispatching and scheduling the AMoD vehicles.	3
FF	Free-flow Network Speed. The free-flow speed or travel times of the city network are based on the speed limit on the network edges.	72
ID	Imbalance Density. The ID links the AMoD supply and demand using reachability density functions.	68
KDE	Kernel Density Estimation. A non-parametric probability density estimation technique often used in statistics.	65
MFM	Mean Factor Method. A heuristic used to scale the network travel times using mean travel time factors.	54
PBD	Prioritized Balanced Density. A lexicographic approach for solving the repositioning problem that first optimizes the supply-demand imbalance density and then the repositioning distance.	99
RD	Reachability Density. A formulation inspired by KDE that links the individual points, either vehicle or customer pickup locations, on the map using reachability functions. In contrast to KDE, each reachability function can have a different bandwidth depending on network speed.	65
RF	Reachability Function. It is a function inspired by the kernel functions used in KDE. In contrast to the KDE, the bandwidths of a reachability function describes the reachable distance from the center of the function within a given travel time limit.	65
RFR	Reachability Functions based Repositioning method. A class of repositioning methods that use the estimated imbalance density for repositioning decisions.	92

Abbreviation	Description	Page
RFRR	RFR with Regions. A formulation of RFR method based that incorporates a disjoint set of AMoD operating regions. The default RFRR formulation only allows the repositioning of idle vehicles from areas of vehicle surplus to areas of vehicle deficiency.	94
RFRRp	RFRR formulation with positive zone flow. A formulation of RFRR method that also allows the repositioning of idle from regions of vehicle surplus to regions of vehicle surplus in addition to deficiency areas.	98
RFRRf	RFRR formulation with full flow. A formulation of RFRR method that does not have any regional restriction, i.e., idle vehicles can reposition from and to any AMoD region.	98
SSM	Squared Scaling Method. A network travel time scaling method similar to ASM but uses the square of the travel time error as the objective function.	54
VCO	Vehicle Control Optimization. An optimization problem used for assigning vehicles to customers.	42
VCO_{idle}	VCO using idle vehicles. A VCO type where only the idle vehicles are considered for assignment	42
$VCO_{enroute}$	VCO using idle and en-route vehicles. A VCO type where idle and en-route vehicles are considered for assignment.	42

Chapter 1.

Introduction

The last two decades have seen a rise in internet-based sharing platforms. This led to a steady growth in the removal of the traditional "middlemen" for the exchange of goods and services or in some cases replaced them with an equivalent online intermediary platform. The phenomenon that initially started with platforms for simple peer-to-peer sharing of files and trading of goods during the 1990s, has grown so immensely that it introduced its own terms like *sharing-economy* into dictionaries [AVITAL et al., 2014]. The shift towards a peer-to-peer collaborative economy was first noted by GANSKY [2010] and further developed by BOTSMAN and ROGERS [2010]. They noticed a shift into collaborative consumption or sharing economy that is less dependent on traditional hierarchies. They believed that this would not only open the door for new market efficiencies but also reframe many established services. As the number and the outreach of companies sharing the new ecosystem grew, the definitions of various terms used also became clearer. Thus, sharing economy is defined as *"an economic system based on sharing underused assets or services, for free or for a fee, directly from individuals"* [BOTSMAN, 2015]. Some examples of shared economy-based businesses and companies include sharing of underutilized accommodation, e.g. Airbnb, and sharing of empty seats in a vehicle for inter-city rides, e.g. BlaBlaCar. Similarly, on-demand services are defined as *"platforms that directly match customer needs with providers to immediately deliver goods and services"* [BOTSMAN, 2015]. Examples of companies providing on-demand services for goods include Foodpanda, Lieferando and Deskbeers.

Perhaps, the biggest area where the sharing economy has significantly "reframed" the traditional outlook is the transport sector. With the introduction of mobility on-demand (MoD) platforms, it is among the leading sectors of sharing economy revolution. MoD is based on the idea that the mobility of people and goods can be offered as an on-demand service. With a myriad of MoD modes like bike sharing, e-scooter sharing, car-sharing, ride-hailing and others, the transport sector has brought seamless transportation of people with the touch of a button [TIRACHINI, 2019]. Among these MoD modes, ride-hailing services have caused the biggest disruption in the transport sector. The ride-hailing smartphone apps match a traveler or customer to a driver who is willing to drive the traveler to his destination in a privately owned vehicle (POV). This is usually provided by a private-sector mobility service provider (MSP), also known as transportation network company (TNC), that maintains the online platforms (including the smartphone apps). The competitive prices, ease of transport and seamless steps from requesting a ride to the final payment have caused widespread acceptance and large-scale usage of MSPs for regular trips. Especially in areas with bad public transport (PT) system, these services represent a viable alternative to POVs [STOCKER and S. SHAHEEN, 2019].

The widespread popularity and success of MSPs are also indicated by the sheer number of customers and covered cities over the years. According to one ranking, three (Uber, Lyft, and DiDi Chuxing) out of the top five most market disrupter companies in 2018 were MSPs [CNBC, 2018]. Uber, launched in 2010, was operating in 400 cities by April 2016 [Z. LI et al., 2016], and increased to more than 900 cities in 2020 [UBER, 2020]. Similarly, DiDi Chuxing, launched in 2012 as a local

ride-hailing service inside China, by 2019 had more than 550 million users and more than 31 million drivers [CNBC, 2019]. It even bought Uber's operation in China in 2016. DiDi, labeling itself as the world's leading mobile transportation platform, even expanded its operation to Mexico, Australia, and Japan [CNBC, 2019].

In addition to MoD services, the autonomous vehicle (AV) technology also got significant momentum in the last decade. The AVs are expected to produce the next paradigm shift in the transport sector. Without the requirement of a driver and more intensive vehicle utilization through central controllers, the autonomous mobility on-demand (AMoD) services are imagined to be cheaper, more efficient, and more environment-friendly [Litman, 2023; Pavone, 2015; Dandl, 2022]. Some companies like Waymoo and Cruise are already in their pilot phase in multiple cities and other companies like Zoox and Moia are also expected to follow in coming years. In anticipation of these developments, several researchers in the last decade have tried to evaluate the benefits and impacts of introducing AMoD in different parts of the world, and potentially develop control approaches that would enhance their advantages. The current dissertation is also a step forward in the same direction.

1.1. Impacts and Variants of Ride-Hailing Services

The usage of POVs and modern communication technologies make the modern MoD services unprecedentedly different from the traditional road-side taxi-hailing services. They provide users with a reliable service, shorter waiting times, and access to larger areas than the PT in many cities — all at a lower price [Rodier and Michaels, 2019]. Several studies have reported that major contributing factors for preferring MoD services, especially ride-hailing services, over other modes of transfer are the ease of payment, shorter waiting times, and the fastest way to reach destination [Rayle et al., 2016; Rodier and Michaels, 2019]. This provided the low-income households, who are not able to own or lease a POV, an opportunity to afford regular trips in comfortable MoD vehicles. The same group of people also disproportionately represents the high-frequency users of MSPs [Lazarus et al., 2021].

Overall MoD services have revolutionized the transport sector, bringing comfortable and affordable transport options to many in society. However, the realization of the ideal imagining of smart mobility — where everyone is visioned to have access to personalized MoD service with clean, efficient, and flexible transport on the tip of their finger [Wockatz and Schartau, 2015] — is highly dependent on how the state and government policies maneuver the transition into new transportation models to make them beneficial to both society and the environment [Docherty et al., 2018; Tirachini, 2019].

Thus, the topic of sustainability and the impacts of MoD services, more specifically ride-hailing, on travel behavior is still under research [Rodier and Michaels, 2019; Tirachini, 2019]. Usually, one of the most important parameters used to measure this impact is the increase in vehicle kilometers travelled (VKT), which can cause an increase in congestion, crashes, noise, and gas emissions. Multiple researches have concluded a likely increase in the overall VKT due to ride-hailing services [Tirachini, 2019]. For example, Henao [2017] reports an 85% increase in VKT due to MSPs in the Denver area. Schaller [2018a] and Schaller [2018b] reports an additional 976 million miles of driving added to New York City (NYC) streets between 2013 to 2017 due to MSPs. He estimates that MSPs have added almost 5.7 billion miles in nine major cities of the United States[1].

A number of approaches are usually suggested to mitigate the negative impacts of increased VKT, ranging from shifts in governmental policies to improving the technologies and methods underlying

---

[1]Boston, Chicago, Los Angeles, Miami, New York, Philadelphia, San Francisco, Seattle, and Washington DC metro areas

the actual services. The sustainability of MSP is dependent on the question of whether the MSP are complimenting or substituting the traditional transportation networks, especially PT [TIRACHINI, 2019]. Policies encouraging the complementary application of MSPs can improve the overall situation. There are already examples of such agreements with varying degrees of success in North America [TRANSIT, 2019]. These projects offer special discounts for going to specific regions or riding to or from train stations and bus stops. The most successful of these projects are those for the workforce and students, providing the trip's first and last mile. This increases the overall accessibility and availability of transport services for the people. J. IACOBUCCI et al. [2017] suggests that the seamless payment technology for all the involved partners is one of the primary reasons for the success of these partnerships.

Besides the changes in the overall policy, improving the methods and techniques used by MSPs can also enhance the overall efficiency. One of these techniques combines the spatially and temporally overlapping customers into single trips, i.e., ride-sharing (RS) (also known as ride splitting). The widespread usage of global positioning system (GPS) equipped cellphones makes it possible to dynamically combine multiple customers into a single ride to increase overall vehicle utilization. Several studies have suggested that RS can cause a significant reduction of required vehicles and VKTs [ALONSO-MORA, SAMARANAYAKE, et al., 2017; SUN and L. ZHANG, 2018; NARAYANAN, CHANIOTAKIS, et al., 2020].

Despite the optimistic simulation results of RS services, in reality, the number of people willing to share rides in current MoD services are limited due to multiple socioeconomic factors [LAZARUS et al., 2021; ALONSO-GONZÁLEZ et al., 2021]. Additionally, with RS services, the differences between a PT and a service offered by private sector MSPs gets even more blurred. The potential worry of RS is that a significant number of customers may shift from PT to RS [TIRACHINI, 2019]. LEWIS and MACKENZIE [2017] studied the short-lived UberHop service in Seattle that combined up to 5 passengers into a ride. They reported that almost 45% of the customers were replacing bus trips, and 79% of the trips were made by a single passenger. Similarly, one study observed that for MSPs (like Uber and Lyft) providing service options of both with- and without shared-ride, the shared rides do not contribute a significant percentage to the total number of served requests [HENAO and MARSHALL, 2019]. Thus, great research potential is available on improving the RS services and incorporating them with the PT to complement each other and enhance the whole system. The future AMoD services can also solve some of the above issues, as discussed in the next section. Nevertheless, so far, the most popular MoD service model among users is serving only a single customer at a time which is also the primary focus of the dissertation. For consistency, the dissertation reserves the term ride-sharing (RS) for the former and ride-hailing (RH) for the latter MoD mode.

1.2. Autonomous Mobility-on-Demand (AMoD) Services

The convergence of electric drive, AV technology, and shared mobility in AMoD services has the potential to transform the transport sector [S. A. SHAHEEN et al., 2020]. Without the requirement of a human driver, the AMoD services are expected to be cheaper than the current MoD and Taxi services due to lower maintenance costs [CHEN et al., 2016; BAUER et al., 2018; NARAYANAN, CHANIOTAKIS, et al., 2020; LITMAN, 2023]. Additionally, the ability to control the entire fleet using a central fleet controller (FC) could significantly improve performance as well as chances of pooling rides [S. A. SHAHEEN et al., 2020; S. A. SHAHEEN et al., 2020; ALONSO-MORA, SAMARANAYAKE, et al., 2017; ALONSO-MORA, WALLAR, et al., 2017]. AMoD services are also expected to have smaller land use primarily due to higher AMoD fleet utilization and reduced private vehicle ownership [W. ZHANG and GUHATHAKURTA, 2017; KONDOR et al., 2019].
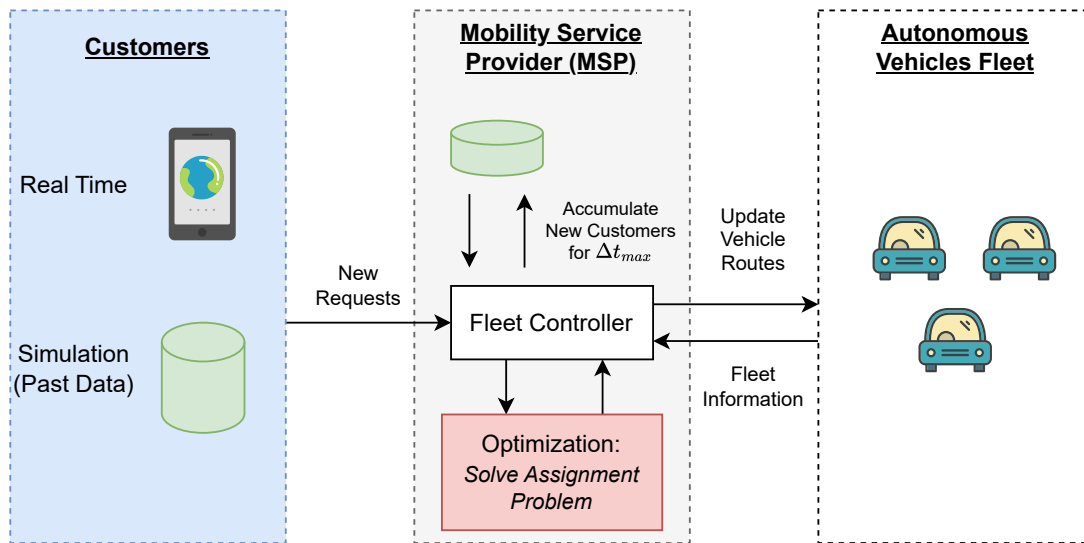
Figure 1.1.: The operation of AMoD services.

Considering the above benefits, AMoD services have a significant potential to solve some of the issues that current MoD services face. A major advantage of AMoD system is that the MSP can fully control AV fleet via central FC. Compared to current MoD systems, the FC can be designed in a way that minimizes the adverse effects by lowering the deadheading (empty VKT without on-board customers) and increasing the vehicle utilization. Alternatively, since there is still some time till AMoD services are widely available, the positive influence of AMoD services on the transport sector can be proactively enhanced via governmental policies and regulations. Similar to the current MoD services, they also have the potential to increase VKT and cause shifting away of the travelers from PT towards AMoD service. This can cause additional congestion in certain parts of the city, especially where short trips are replaced by AMoD trips [Moavenzadeh and S. Lang, 2018]. These can be avoided by adopting suitable strategies for their wide-scale adoption [J. Iacobucci et al., 2017; Stocker and S. Shaheen, 2019].

In view of the above, there is an active requirement to optimize the performance of AMoD services. The current dissertation is also an attempt in this direction — improve the efficiency of the AMoD services so that more customers could be served with a smaller fleet and lower VKT. Instead of finding optimal prices and governmental policies, the dissertation focuses on innovative procedures for a better assignment of vehicles to customers and the maintenance of the AMoD fleet. The dissertation focuses mainly on improving RH based AMoD services. The following section briefly presents the overall handling of customers and the fleet in the AMoD services, based on which the dissertation will look deeper into the areas of potential improvements.

Operation of Autonomous Fleet Ride-Hailing Service

The AMoD services in the dissertation consist of shared autonomous vehicles (SAVs) that are entirely controlled by a central FC located at the MSP. The major difference between the current MoD and AMoD services is the assignment of customers to vehicles. The MoD services include human drivers that can accept or reject the assigned customers. On the contrary, the central FC fully controls the customer assignment decisions in an AMoD service. Similarly, since the drivers in the current MoD services are mostly individuals with POVs who earn variable salaries depending on the number of

customers served and the distance traveled with customers on board, the MSPs have to take care that the customers and maintenance tasks are fairly assigned. Contrarily, the whole fleet of AVs is assumed to be managed by MSPs; thus, the performance is evaluated for the whole system, even if a group of AVs generates small revenue or completely go into loss.

The rest of the operation of an AMoD service is quite similar to the current MoD services as shown in Figure 1.1. The customer requests a ride using a smartphone app that the MSP needs to serve. Inside the MSP, the new requests are handled by a central FC that solves the vehicle assignment problem and sends the updated paths to the fleet. The FC is also responsible for keeping track of all operations and vehicle states and recording relevant data.

The optimization problem of assigning vehicles to dynamic customer requests comes under the category of stochastic and dynamic vehicle routing problem (SDVRP). These problems are *dynamic* — the new customer requests or locations to be visited are added to the system over time — and *stochastic* — a part of the unknown data or future users are only known in terms of statistical information such as regional forecasts or probability distribution [TOTH and VIGO, 2014, Chapter 11][AGATZ et al., 2012]. Commonly, researchers use a *batching* strategy to deal with the dynamic aspect of the SDVRP problem; the dynamic customer requests are periodically grouped together before solving an optimization problem for assigning vehicles to customers. The dissertation refers to this optimization problem as vehicle control optimization (VCO). The overall goal of the batching method is to reach a better solution to the whole dynamic problem by optimizing small batches of customers. Some MoD services also claim to use a similar batching technique [UBER, 2022].

The dissertation improves the overall AMoD performance by developing and incorporating novel FC methods in its operation. The next section discusses the core areas of focus in more detail.

1.3. Research Context and Objectives

The development of solution methods for SDVRPs is a vast research area in traffic engineering and operations research, with a myriad of challenges and requirements for each variant of SDVRP. The studied variant of SDVRP — AMoD with RH — is no exception. Improving all potential aspects of AMoD services, i.e., governmental policies, dynamic pricing, vehicle assignments, and others, is beyond the scope of a single dissertation. Therefore, this section presents the core improvement areas focused on in the dissertation.

The overall operation of an AMoD service can be described as a Markov decision process (MDP), where a decision taken at any stage can lead to a totally different outcome. Thus, short-sighted decisions can lead to an overall worse performance. Figure 1.2 shows the primary operational loop of the studied AMoD system. The VCO assigns vehicles to customers in a batch. Usually, the customer origin and destination pattern differs throughout the day, especially when people go and return from work [DANDL, M. HYLAND, et al., 2020]. Thus, short-sighted assignment decisions without considering the long-term impacts on the system lead to a supply-demand imbalance in different parts of the operation area. This means that many vehicles can end up in regions where they will remain idle for a long time without customers.

In literature, the above problem is solved by incorporating statistical information into the AMoD system, which mainly comprises a forecast of system state and customer locations [PAVONE et al., 2012; DANDL, M. HYLAND, et al., 2019; DANDL, M. HYLAND, et al., 2020]. This information is then used in a repositioning strategy, where idle vehicles are regularly repositioned to high demand areas. The dissertation focuses on similar techniques for improving the AMoD services as shown in Figure 1.3. The objective does not include devising an accurate or improved statistical model for predicting potential users; instead, the focus is on its incorporation into the operation of AMoD
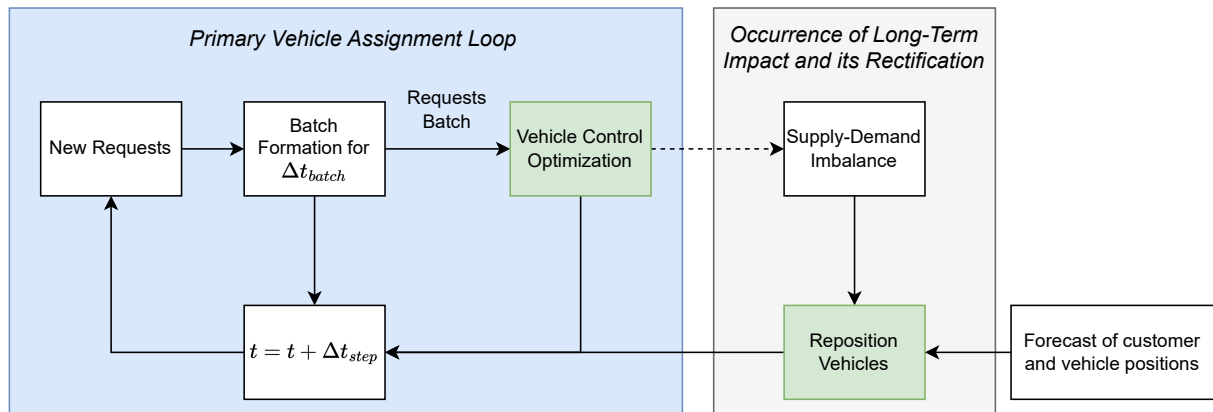
Figure 1.2.: The primary operational loop of AMoD services

system.

Thus, the overarching objective of the dissertation is to answer the following main question:

- **MRQ**: How to improve the overall AMoD performance using novel FC methods?

To answer the above question, the dissertation aims to investigate the following aspects and research questions:

1. **Vehicle Distribution Model**:

   For various maintenance reasons, the AMoD literature often divides the operation area into regions. Most operational algorithms that use regional dissection of the operational area assume individual regions to be independent: the vehicles within a single region can only serve customers within that specific region. However, this is only valid when the assumed regions are relatively big, and the city network and customer demand are such that many trips start and end in the individual regions. Nonetheless, from the perspective of developing operational algorithms, the region sizes (within a city) cannot be increased indefinitely. Large regions would impede the full potential of any algorithm, as vehicles within a region would not be able to pick up customers from all locations while maintaining a short pickup duration.

   As a result, the defined regions are often small and contiguous where the neighboring regions are no longer independent. Therefore, there is a strong requirement for a vehicle distribution model that is loosely bound to regional shapes and could coherently take into account the state of the neighboring regions and the whole operation area.

   Therefore, the thesis studies the following questions:

   - **RQ 1.1**: How can a vehicle distribution model be defined that is independent or at least loosely bound to regional shapes?
   - **RQ 1.2**: Does the above distribution model has a strong and consistent relationship between states of individual regions and the overall AMoD system? How can it be measured?

2. **Repositioning problem**:

   The most common type of fleet maintenance algorithm that utilizes the operational regions is repositioning idle vehicles to demand-intensive regions. It uses statistical information to

periodically reposition idle vehicles to demand-intensive regions. This resolves the supply-demand imbalance accumulated over time. The thesis aims to improve the repositioning algorithms for the AMoD services. The aim is to find a consistent procedure that considers the imbalances of the overall operation area in a continuous 2D domain instead of dealing with individual regions separately. In this regard, the vehicle distribution model developed above is utilized. Thus, the thesis addresses the following question:

- **RQ 2.1**: How can the newly developed vehicle distribution model be incorporated into the repositioning problem?

3. **Vehicle assignment problem**: As the AMoD operational problem is basically an MDP process, the solution of the VCO determines not only which customers are served next but also where the vehicles will end up in the long run. Most of the current VCO methods only consider the short-term profit of the current batch. However, the statistical information can also be incorporated into the VCO to consider the long-term impacts of VCO decisions on supply-demand imbalances. The intuition is to *implicitly* reduce supply-demand imbalance in small temporal steps inside the VCO such that the need for an *explicit* repositioning at a later stage is minimized. This would also guarantee that the positive effects of the repositioning decisions remain valid for extended periods, increasing the overall AMoD performance. This represents an *anticipatory* or *proactive* approach to reduce the long-term supply-demand imbalance. Thus, the dissertation addresses the following questions:

- **RQ 3.1**: How to incorporate the statistical information inside a VCO?

- **RQ 3.2**: How to utilize the vehicle distribution model developed in RQ 1.1 into the proactive VCO?

1.4. Research Methodology and Thesis Structure

Figure 1.3 summarizes the main areas of AMoD operations focused in the dissertation. First, to address RQ 1, the dissertation requires an environment where AMoD performance could be reliably measured. However, the AMoD systems are still in the pilot phase, with several companies like Waymo LLC [ROTH, 2022] and Cruise LLC [HAWKINS, 2021] operating within a limited service area. Without an actual AMoD service or even a fully autonomous vehicle, the dissertation evaluates any AMoD performance improvement in computer simulations. To produce consistent and reliable results, the simulations must be as detailed as possible. Before any improvement for AMoD services is devised, developing a detailed AMoD simulation is crucial. Therefore, chapter 3 builds an agent-based simulation framework where customers, vehicles, and AMoD operator are modeled as agents. Instead of a simplistic 2D environment, these agents interact with each other on an actual city map derived from OpenStreetMaps (OSM) [OPENSTREETMAP, 2017]. Typically, the OSM map only contains free-flow network travel times based on the speed limits of individual roads. The free-flow speeds may produce unreliable AMoD performance as the simulated AVs can travel much faster on the map than in reality. Thus, chapter 3 also investigates how to scale the travel times of each road link such that the overall network speed is as close as possible to reality. For this purpose, chapter 3 introduces a novel spatiotemporal scaling method for the city network. The second most important requirement for a successful AMoD simulation is the generation of realistic customer demands. Thus, instead of randomly generating customer requests, the dissertation uses the open-source NYC taxi data set [*TLC Trip Record Data - TLC* 2023]. This guarantees that the customers represent actual travel demand in the city, and any improvement observed for AMoD system is not due to unrealistic
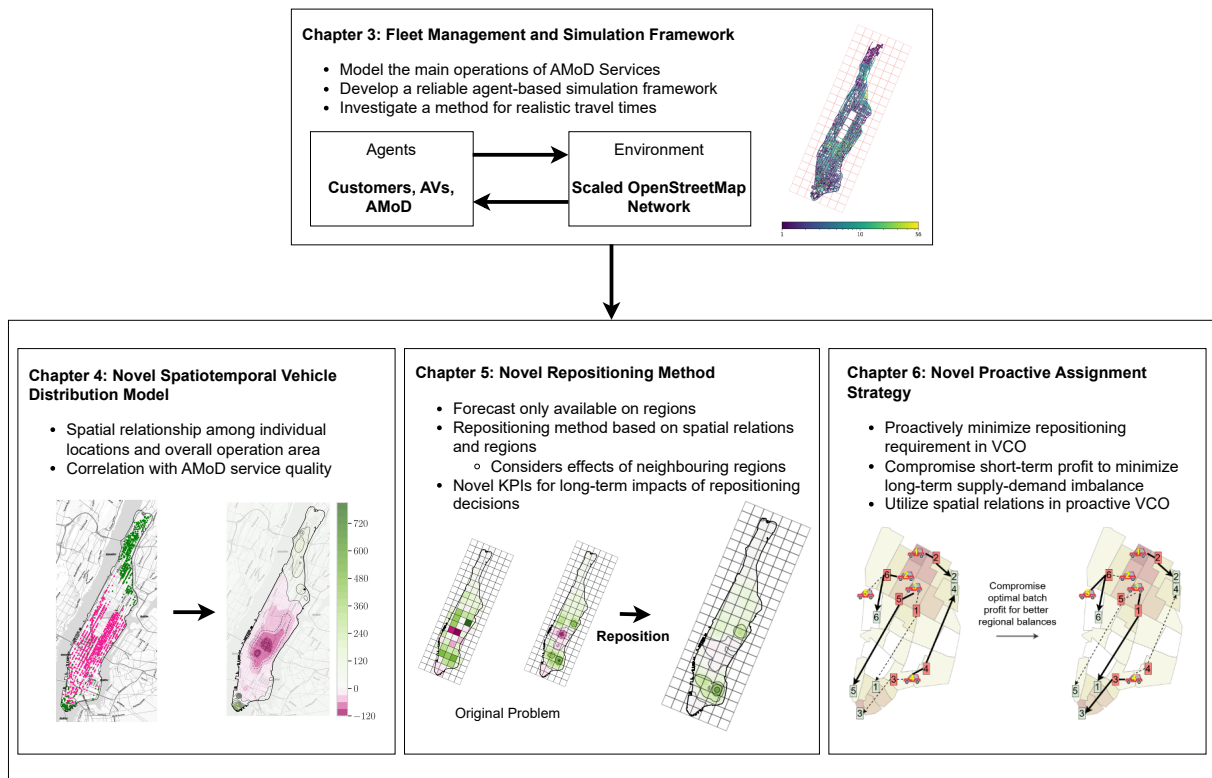
Figure 1.3.: Overview of the research methodology

travel demands. The simulation framework built in chapter 3 is used throughout the dissertation to measure the efficiency of the introduced FC methods.

Chapter 4 later focuses on RQs 1.1 and 1.2. The chapter primarily aims to develop a spatial relationship that directly deals with individual coordinates while simultaneously taking into account their surrounding locations. This removes the requirement of defining regions for the aggregation of geographical locations. The chapter measures the effectiveness of the developed spatial relations by investigating how well these relations correlate with AMoD service quality offered to customers. If the service quality provided at a customer pickup location could be distinguished based on these spatial relations, then the developed spatial relations provide an effective alternative to aggregation over regions. This would also entail that any FC method that actively uses these relations for AMoD fleet management will also improve the overall AMoD performance.

Afterwards, chapter 5 builds on top of chapter 4 to answer RQ 2.1; it actively uses the above mentioned spatial relations to incorporate forecast information into the repositioning step. It develops a novel repositioning method that uses these spatial relations inside the repositioning formulation. Since the spatial relations consider the surrounding area in its formulation, the newly developed repositioning formulation inherently considers the impacts of multiple regions in the repositioning decisions. This is in contrast to the traditional repositioning strategies that assume independent regions without considering the effects of repositioning decisions on the surrounding areas.

Later, chapter 6 focuses on RQ 3.1 and 3.2. It first uses the spatial relations developed in chapter 4 to study why VCO causes supply-demand imbalances in different parts of the AMoD operation area. Then it investigates how this can be avoided by making better decisions in the VCO. It develops a proactive VCO formulation where each vehicle assignment decision is made keeping in

view the long-term impacts on vehicle supply-demand imbalances. Primarily, it assigns the vehicles to customers in two steps. The first step solves the usual VCO and calculates the optimal value of the objective function. The second step compromises a percentage of the optimal objective function value of the first step to achieve a VCO solution that potentially minimizes the supply-demand imbalance in the long run. It uses the spatial relations built in chapter 4 to measure the long-term impacts. Additionally, it further explores the performance of proactive VCO when combined with the repositioning method developed in chapter 5.

Finally, the dissertation discusses the overall outcome of the research in the last chapter. The chapter also summarizes the assumptions followed throughout the dissertation and the potential limitations due to these suppositions. It additionally discusses the possible areas of improvement of the current work.

Chapter 2.

Background and Literature Review

This chapter presents the necessary background knowledge required to understand operational problems in AMoD services and discusses the AMoD studies found in the literature. It first starts with describing the information flow in the AMoD operation. This helps the reader to identify the underlying basic theory and methods used when modeling an AMoD service. Then, the later sections briefly describe the background knowledge and mathematics involved in these models; these include mathematical optimization and its types (section 2.2), integer programming problem (IP) and its commonly used solution methods (section 2.3), and graph theory along with some problems related to AMoD operation (section 2.4). After presenting the necessary background knowledge, section 2.5 reviews the AMoD service types and modeling techniques found in the literature.

2.1. AMoD Services Operation

Before reviewing the AMoD studies available in the literature, it is important to briefly understand the methods and techniques used by the dissertation to model the AMoD operations. This will help the reader to clearly place the AMoD modeling procedure used in the dissertation among the myriad of modeling possibilities found in the literature. Additionally, this section describes the primary technique underlying each of the AMoD components, mainly consisting of mathematical optimization, graph theory, and some typical problems over graphs. Even though this section mentions these underlying techniques from the dissertation perspective, they are the method of choice for modeling many AMoD components in literature.

The dissertation uses an agent-based simulation framework to evaluate the AMoD services. The main agents in the simulation are customers, AVs, and the AMoD operator. These agents interact with each other in an environment mainly consisting of the underlying city network. In the previous chapter, Figure 1.2 already described the main time loop of the AMoD operation. In contrast, Figure 2.1 describes the flow of information and the underlying technique used to model the whole AMoD operation. It should be noted that Figure 2.1 describes the AMoD operation as it is modeled in the dissertation; the AMoD studies in literature can model them with slight differences, as will be discussed in the literature review (section 2.5.4).

The underlying city network is modeled as a graph with nodes representing the intersections and edges representing the roads connecting those intersections. Besides the travel distance and travel time, each network edge also contains an additional attribute for the exact geometry of the road. The customers and AVs are located on the nodes of the city network. Additionally, an AV traverses the edges of the city network for moving from one location to another. In each simulation step, the AV and customer states are updated and sent to the AMoD operator. After performing the internal AMoD fleet management tasks, the AMoD operator sends AMoD offers and updated paths to customers and AVs, respectively.

The dissertation considers two main tasks inside the AMoD operator: VCO and repositioning. VCO
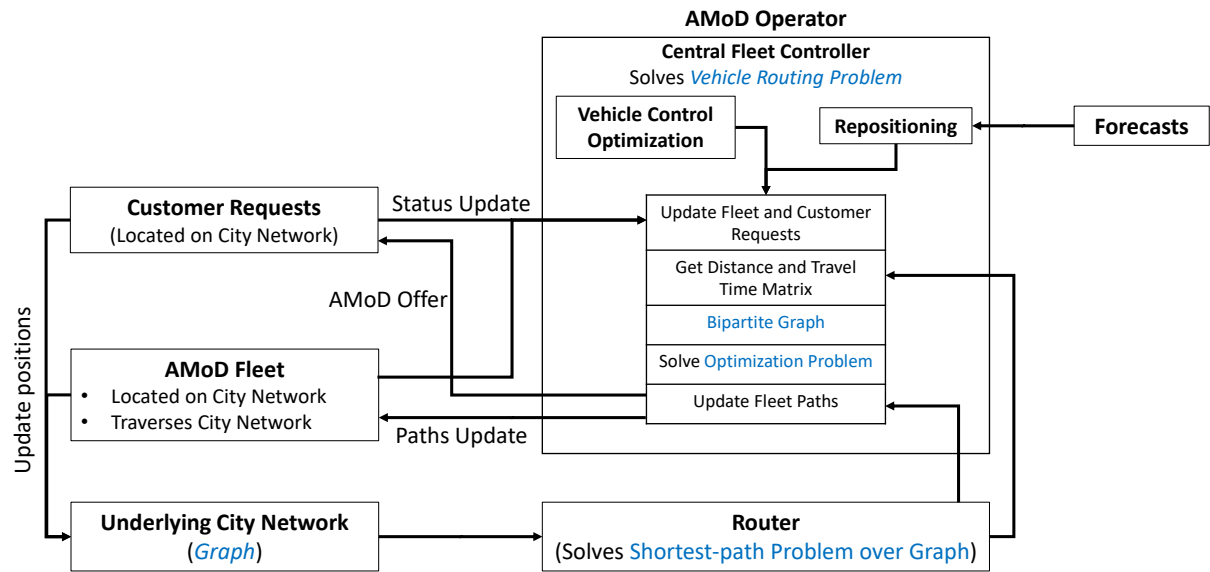
Figure 2.1.: The main elements of AMoD operation. The underlying techniques used to model them are marked with blue color.

assigns AVs to customers while repositioning sends idle AVs to areas of higher customer demand. However, both operations mainly consist of similar steps. After updating the customer and AV states, all the necessary information is gathered for forming an optimization problem. In this regard, the first thing required is the distance and travel time matrices for all the combinations of locations. This is provided by the router, which solves the shortest-path problem over the city network graph. Second, a bipartite graph is formed. For VCO, the bipartite graph is between AVs and customers[1], while the bipartite graph is between idle vehicles and destination locations for repositioning. Third, an optimization problem is formed to solve the bipartite matching problem; the solution provides the destinations in the path of each AV. Depending on the formulation, the optimization problem can belong to any of the subcategories of mathematical optimization (section 2.2.2 provides relevant classifications). At this stage, the AMoD asks for the exact routes for each origin-destination pair of the modified AV paths. The exact routes consist of detailed information on nodes and edges between each origin-destination pair. These updated paths are sent to AVs. Similarly, the customers are informed of the assignment decisions.

2.2. Mathematical Optimization

Mathematical Optimization is a widely used sub-field of mathematics. It is often also called *Mathematical Programming*, *Numerical Optimization*, or simply *Optimization*. Generally, it is the science of finding the best solutions to mathematically defined problems that may arise in any quantitative discipline [SNYMAN and WILKE, 2018; NOCEDAL and WRIGHT, 2006]. The formulated problems may model a physical reality or a management system. The former generally relate to Science and Engineering, where minimum energy configurations are often required for general structures, ranging from molecules and electrical circuits to suspension bridges and aircraft dynamics. The latter generally relate to societal and industrial applications where commercial and financial decisions need to be

---

[1]some studies in literature form a bipartite graph between AVs and trips. A trip can consist of multiple customer and repositioning locations
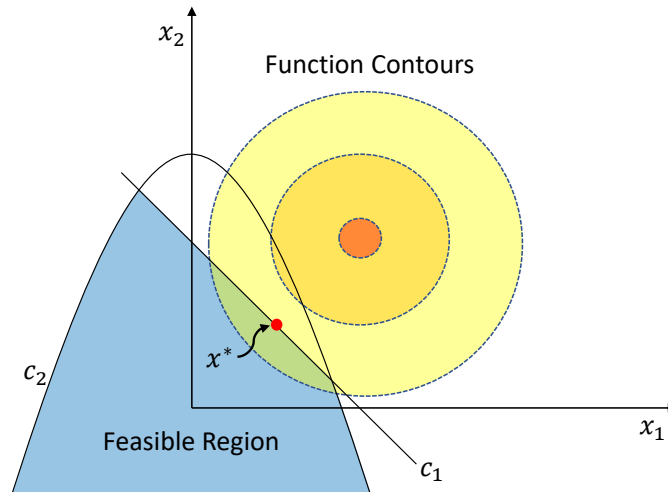
Figure 2.2.: An example of 2D linear optimization (minimization) problem

taken that maximize or minimize certain parameters like profit or cost, respectively. The dissertation topic also falls into this category where the goal is to maximize the MSP profit while maintaining a certain service quality.

This section briefly overviews the optimization problems and common solution methodologies. The following descriptions are based on the standard books on the topic [SNYMAN and WILKE, 2018; MATOUSEK and GÄRTNER, 2007; NOCEDAL and WRIGHT, 2006; KORTE and VYGEN, 2018].

2.2.1. Mathematical Formulation

In mathematical terms, optimization is the *maximization* or *minimization* of a function while satisfying constraints on its variables. As such, it consists of the following:

- A vector $\boldsymbol{x} \in \mathbb{R}^n$ of design or decision variables

- An objective function $f : \boldsymbol{x} \mapsto \mathbb{R}$ to be maximized or minimized

- Set of functions $g : x \mapsto \mathbb{R}$ and $h : x \mapsto \mathbb{R}$ that define the equation and inequality constraints that the variables in $\boldsymbol{x}$ must satisfy, respectively.

With the above notations, an optimization problem is written as:

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) \tag{2.1a}$$

$$\text{subject to} \quad g(\boldsymbol{x}) = 0 \tag{2.1b}$$

$$h(\boldsymbol{x}) \geq 0 \tag{2.1c}$$

Eq. 2.1 describes a minimization problem; alternately, the problem would be a maximization problem if the objective was to maximize $f(\boldsymbol{x})$. The constraints bound the space into areas of possible solutions or set of points satisfying all constraints, which are called as *feasible set* or *feasible region*. Any point in this set represents a *feasible solution*. Figure 2.2 shows an example of a feasible region.
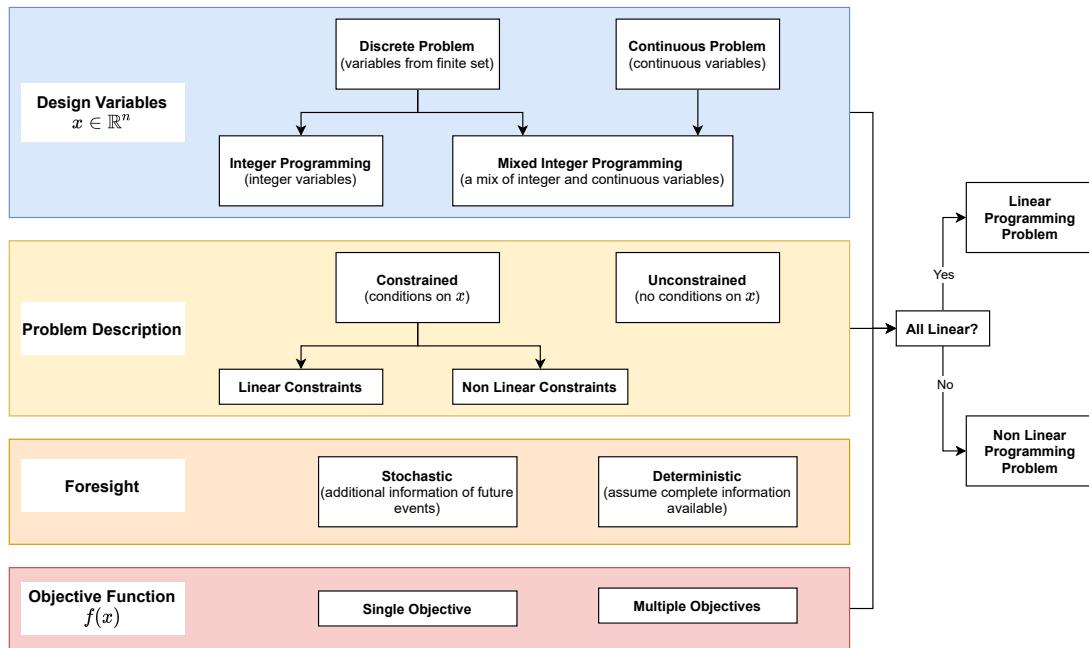
Figure 2.3.: Classification of optimization problems.

2.2.2. Classification of Optimization Problems

After the general mathematical formulation of an optimization problem, this section introduces their classification. Since the dissertation uses some specific types of optimization problems, it is important to understand their differences briefly. There are multiple ways in which an optimization problem can be classified. The following only presents a basic classification based on the nature of the objective function, used constraints, and the design variables.

Discrete and Continuous Optimization Problems

Eq. 2.1 presented the general optimization problem with continuous variables, i.e., $\boldsymbol{x} \in \mathbb{R}^n$. However, a lot of times, the variables can only take integer values. For example, if a company wants to optimize the number of products produced for a specific season or the number of different plant types required to fulfill the customer demand in a region. In that case, fractional values of design variables do not make sense. In such cases, for some of the variables, the mathematical formulation has *integrality constraints*, i.e., $x_i \in \mathbb{Z}$, or *binary constraints*, i.e., $x_i \in \{0, 1\}$. The latter is used for decisions involving only *Yes* or *No* answers. If all variables are integer or binary, the optimization problem is known as *IP* problem. In contrast, if only some of the variables are integer or binary, then the problem is called a *mixed integer programming problem (MIP)* problem. In the last few decades, the terms IP and MIP are also being used interchangeably.

IP problems belong to a bigger class of optimization problems known as *discrete optimization* problems which are not limited to only integer and binary variables, but they can contain more abstract variables like permutations of an ordered set. The distinguishing feature of these problems is that the variables are drawn from a finite set. In contrast, the feasible set of *continuous* problems consists of the infinite set as the variables are real numbers, and the smoothness of the objective function allows us to derive the potential behavior of the objective function and constraints in the neighborhood of a point. Contrarily, the behavior of a discrete problem can change significantly in

the neighborhood of a point, making discrete problems much more difficult to solve than continuous ones. In mathematical terms, the discrete problems show an extreme form of non-convex feasible sets.

Constrained and Unconstrained Optimization Problem

If the optimization problem has no constraints on the solution space, it is called an *unconstrained* optimization problem; otherwise, it is called *constrained*. These constraints arise due to some external limitation, for example, the available budget for producing goods or, in the case of this dissertation, the maximum waiting time an MSP customer is allowed to wait.

The constraints can comprise of *linear* or *non-linear* relationship among design variables. If both constants and the objective functions are linear, the problem is called *linear programming* problem. Linear problems are the most commonly modeled and solved optimization problems. In contrast, if some of the variables or the objective function are non-linear, the problem is known as *non-linear programming* problem, which is more challenging to solve than a linear problem.

Stochastic and Deterministic Problems

Sometimes an optimization problem cannot be fully specified at the time of formulation. The optimal system behavior might depend on a future quantity that is unknown at the time of solving the problem. Such characteristics can appear in economic or financial planning models; for example, the exact demand for a new product might not be known during production. Another example is the greedy, short-sighted assignments of vehicles to customers that might lead to long-term system imbalances. In these cases, the modeler prefers to incorporate additional knowledge instead of making decisions based on the information already present. For example, a statistical model of the most probable scenarios under which specific demand might arise. Such optimization problems are referred to as *stochastic optimization* problems. Alternatively, the problem is called a *deterministic optimization* problem if such considerations are not taken into account.

Single and Multi-Objective Optimization

If the optimization problem consists of multiple objective functions instead of a single objective, then the problem is known as *multi-objective optimization* problem. Ideally, these problems try to minimize (or maximize) all objectives. But, most such problems consist of *conflicting* objective functions, i.e., if one objective is minimized, then the other objective increases and vice versa. Therefore, multiple *non-dominated* Pareto optimal solutions are usually sought where none of the objectives can be improved without degrading other objective functions. Note that the classifications described in the previous sections also apply to multi-objective problems.

2.3. Integer Programming Problems

The previous section introduced mathematical optimization and its different problem types. Since the dissertation mainly formulates the AMoD operation as an IP problem, this section briefly describes IP problems and their solution methods in more detail.

As described in the previous section, IP problems consist of optimization problems in which all the decision variables are integers. The objective function and constraints in these problems could be either linear or non-linear functions.

Mathematical Formulation

The *linear* and *quadratic* IP can be described in the following matrix form without loss of generality:

$$\min_{\boldsymbol{x}} \quad \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} \tag{2.2a}$$

$$\text{subject to} \quad \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \tag{2.2b}$$

$$\boldsymbol{x} \in \mathbb{Z}^n \tag{2.2c}$$

where $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ is a symmetric matrix and $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is the matrix for m constraints, and $\boldsymbol{c} \in \mathbb{R}^n$ is the cost vector. Eq. 2.2 describes an integer quadratic programming problem (IQP). If the matrix Q is a null matrix, the problem becomes an integer linear programming problem (ILP). Similarly, if the integer constraints in Eq. 2.2c are removed, then the problem becomes a *continuous* quadratic programming problem (QP) or linear programming problem (LP) (with null $\boldsymbol{Q}$).

Solution Methodologies

Generally, IP belong to a class of problems that are computationally difficult to solve, known as NP-hard problems [Korte and Vygen, 2018; Conforti et al., 2014]. Solving ILP and IQP is considered an NP-hard problem, making them difficult to solve for large constraints and variable sizes. Practically, it means that if a solution is provided, it takes polynomial time to verify its feasibility. However, the time required to find the optimal solution, by exploring the whole solution space, increases exponentially. It should be noted that the ILP and IQP are NP-hard as a *group of problems*, meaning that some particular problems in this group could still be solvable in polynomial time. For example, the assignment problem is a polynomial time solvable ILP using the Hungarian method with complexity $\mathcal{O}(n^3)$. In contrast to ILP and IQP, all problems in the group LP are solvable in polynomial time using the simplex method.

There are multiple *exact* methods available for solving the IP. Exact methods try to find the optimal solutions along with mathematical surety that the found solution is optimal. However, for larger instances, sometimes the solution space becomes intractable and the computation time required by the exact methods becomes impractical. Therefore, the literature also mentions several *heuristics* and *meta-heuristics* that try to find a *good* sub-optimal solution within practical computation time. *Meta-heuristics* is also a form of heuristic that groups together several heuristics inside a single procedure. They try to avoid getting stuck at local minima (or maxima) in search of better solutions, for which they sometimes accept worse than the current solution.

For the IPs, a commonly used solution approach is to *relax* certain constraints that are easy to solve and provide a good estimate of the original problem. The *relaxations* and the associated problems are solved multiple times till an optimal solution to the original problem is found. The most common among these relaxations is the *linear relaxation*. In this method, the integer constraints of the IP are relaxed, and the obtained LPs is solved multiple times. The following briefly describes the method to solve these relaxed LPs.

These LPs are typically solved using the well-known *simplex* method. In the simplest terms, the simplex method moves from one endpoint of the feasible solution space — *basic feasible solution*— to another till the optimal solution is found. The reader is referred to [Matousek and Gärtner, 2007] for more details on the mathematics involved in the simplex method. The simplex method is used within another method known as *branch-and-bound (B&B)* method for solving an IP. B&B is a systematic search method to find the optimal solution to discrete optimization problems. In its core functionality, it splits up the search space into smaller spaces and keeps track of the best and current

solutions to narrow down the search space further. It solves a series of LPs obtained by relaxing the integrality constraints and limiting the search space with additional constraints.

In the B&B method, determining an efficient pruning of the search tree plays a significant role. For an efficient pruning of the search space, the problem is formulated in a way that the gap between the optimum value of the ILP and the solution of LP is reduced — referred to as *tightening* the formulation. A tighter bound of the objective function provides earlier pruning of the branches of B&B tree. A popular way of doing that is using *cutting plane* method. Similar to B&B, this method solves an LP obtained by relaxing the integrality constraint of the ILP. However, instead of branching over variables, new linear inequality constraints are recursively added till the optimal integer solution is found. The added constraint separates the current solution of LP from the actual feasible set, and thus, referred to as *cut*. An even more popular approach for solving IPs is the *branch-and-cut (B&C)* method which combines B&B and the cutting plane method. It was first introduced by Padberg and Rinaldi in 1987 [PADBERG and RINALDI, 1987], and since then, it remained the most successful method for solving IP. Most of the commercial MIP solvers are also based on B&C method.

The dissertation uses the commercial solver named Gurobi for the above mentioned methods. It uses multiple algorithms, among which *simplex* method is one of the two major algorithms used. Gurobi provides a possibility to model several optimization models, including both ILP and IQP, and has interfaces to multiple programming languages.

2.4. Graph Theory and Relevant Graph Problems

Graphs provide a fundamental structure used for various optimization problems. They provide a systemic procedure to describe relationships among various finite sets. They are extensively used in the dissertation to model the optimization problems in AMoD services. Therefore, the following briefly presents the fundamentals of graph theory followed by graph problems related to the dissertation.

In the most general sense of the term, a *graph* G consists of a finite set of *nodes* or *vertices* V with some (or all) pairs of *nodes* connected to each other via a finite set of connections known as *edges* E. The two main categories of graphs are *undirected* and *directed* graphs.

Definition 1 *An **undirected** graph G consists of an ordered pair $G = (V, E)$ where V is a finite set of nodes and $E \subseteq \{\{v, w\} \mid v, w \in V, v \neq w\}$ is an **unordered pairs** of nodes representing the graph edges.*

Definition 2 *A **directed** graph or **digraph** G consists of an ordered pair $G = (V, E)$ where V is a finite set of nodes and $E \subseteq \{\{v, w\} \mid v, w \in V \times V, v \neq w\}$ is an **ordered** pairs of nodes representing the graph edges. The edges of a directed graph are sometimes called directed edges or arcs.*

The above definitions only allow edges between two distinct nodes. Such edges are said to be *simple* edges. In contrast, if multiple edges connect the same pair of nodes, then the edges are said to be *parallel*. For the dissertation, only the *simple* graphs are relevant which by definition do not have any parallel edges.

Many times all edges in the graph are not equally important for a particular purpose. For example, while finding the shortest path through a city network, each edge may represent different path lengths. In such cases, weighted graphs are used.

Definition 3 *A **weighted** graph has a number or weight assigned to each edge. The weight can represent different variables depending on the problem modeled, for example, cost, capacity, distance, length, or time.*

Bipartite graphs and their *matching* play an important role in the dissertation. They are especially relevant when two clearly distinct groups of nodes exist in the graph. In the later chapters, the dissertation will use *bipartite* graphs for the assignment and repositioning of idle vehicles.

Definition 4 *A **bipartition** of an undirected graph means that the nodes in a graph $G = (V, E)$ can be split into two disjoint sets A and B such that (1) $V = A \cup B$ and (2) all edges have one endpoint in A and the other endpoint in B. Such a graph is known as **bipartite** graph.*

Definition 5 *A set of pairwise disjoint edges $\hat{E} \subseteq E$ (i.e. edges without any common endpoints) in an undirected graph $G = (V, E)$ is called as **matching**. A node is considered **matched** if it is an endpoint of one of the edges in $\hat{E}$.*

Definition 6 *A matching containing the largest possible number of edges is known as **maximum** matching.*

Definition 7 *A matching containing the largest possible number of edges is known as **maximum** matching.*

After presenting the graph theory fundamentals, the following briefly presents the most important graph theory problems related to the dissertation.

Assignment Problem

In a weighted graph, matching with the largest (smallest) possible sum of the weight edges is called **maximum (minimum) weight matching**. In a weighted bipartite graph, the maximum (or minimum) weight matching problem is also referred to as the **assignment problem**. The assignment problem is a fundamental combinatorial optimization problem. Figure 2.4a shows an example of the assignment problem. It is a commonly used problem in AMoD literature. The dissertation also uses it to solve the operational problems in AMoD services.

The assignment problem can be optimally solved using the Hungarian method [KUHN, 1955]. Alternatively, due to the unimodular constraint matrix, the integrality constraint can be relaxed and the problem can be solved using linear programming. The dissertation uses the latter approach.

Shortest Path Problem

The shortest path problem is one of the most widely used graph optimization problems. Given a weighted digraph $G = (V, E)$ and nodes $s, t \in V$, it consists of finding the shortest path P between s and t that has the least possible sum of edge weights. If no such path exists, then the problem decides that t is unreachable from s. It is widely used in real-world applications such as smartphones and navigation devices to find the best route. Figure 2.4b shows an example of the shortest path problem.

There are several variants of the shortest path problem depending on whether the given graph is directed, has edges with negative weights, or the shortest path is sought for all ordered pairs of nodes. Several algorithms exist in the literature for each of these variants, for example, Dijkstra's algorithm (for single source and non-negative edge weights) [DIJKSTRA, 1959], Bellman-Ford algorithm (for single source and general edge weights) [BELLMAN, 1958; FORD, 1956], Floyd–Warshall algorithm (for all pairs of nodes) [FLOYD, 1962; WARSHALL, 1962].

Dijkstra's algorithm has multiple variants which are often used by routing engines for road networks. The shortest path problem on the road networks plays an important role in the operation of the

(a) Assignment problem

(b) Shortest-path problem

(c) TSP example

(d) VRP example with two vehicles
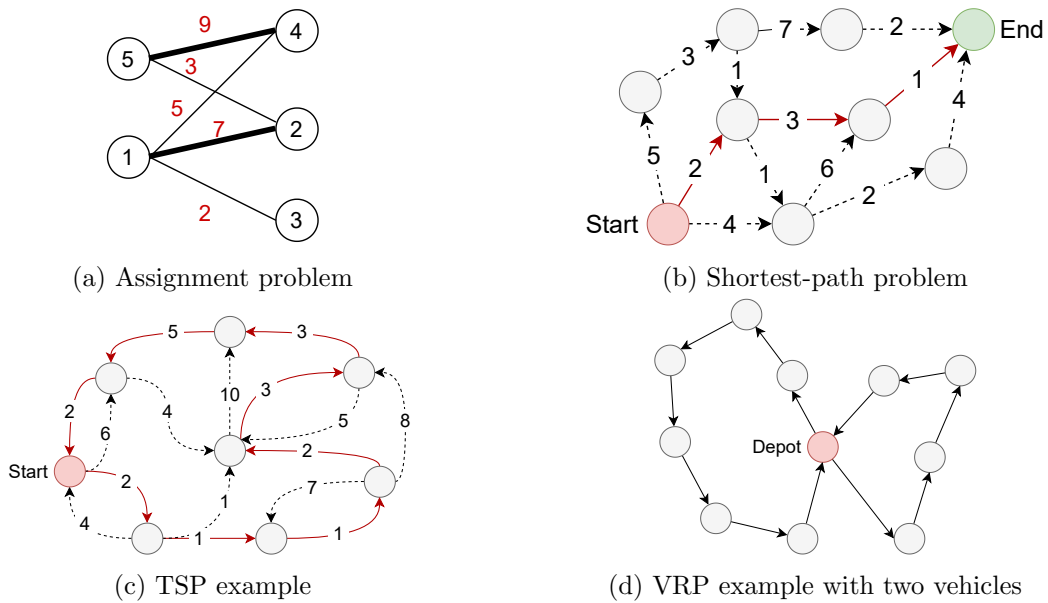
Figure 2.4.: Main graph problems related to the dissertation.

AMoD services. It provides the route followed by AMoD vehicle between two geographical locations. Additionally, for efficiently assigning customers to vehicles, any algorithm requires the expected travel times and distances between customer and vehicle locations. Since the dissertation does not primarily concerns with improving the shortest path algorithm, it simply uses Dijkstra's algorithm as explained in later chapters.

The Travelling Salesman Problem

The travelling salesman problem (TSP) is a well-known NP-Hard problem. It consists of a traveling salesman that must visit n cities with the minimum traveling cost and return to the city of origin. It is usually modeled as a weighted graph with n nodes. The graph can be directed (for asymmetric TSP) or undirected (for symmetric TSP). Each edge weight represents the traveling cost between the two cities. The solution must visit each node of the graph exactly once such that the sum of total edge weights is minimum. Figure 2.4c shows an example of TSP.

The Family of Vehicle Routing Problems

The vehicle routing problem (VRP) is a generalization of TSP where a fleet of vehicles must serve a set of customers with the minimum traveling costs. There are multiple variants of VRP, known as the family of VRPs. In generic terms, the family of VRPs consists of a fleet of vehicles and a set of *transportation requests*. These problems build a set of vehicle paths or routes that serve all or part of the transportation requests at the minimum possible cost while fulfilling all constraints. The solution consists of the information on which vehicles will serve which requests and in which order [TOTH and VIGO, 2014]. Figure 2.4d shows an example of the classical VRP.

In the VRP, transportation requests are associated with geographical points, i.e., locations where the transportation service needs to be provided. VRPs are often modeled as graph problems with the transportation requests as vertices. In a VRP, the edges usually represent the shortest or cheapest path between the nodes with each edge resulting from multiple street segments. The traveling time

or distances in a VRP can be calculated using the routing element of a geographical information system (GIS) [K.-T. CHANG, 2006], the modern examples of which are Google Maps[2] and OSM[3]. The routing element in a GIS solves the shortest path problem (section 2.4) using constant or real-time information. The basic AMoD fleet management problem is a type of VRP as will be discussed in section 2.5.4

2.5. Literature Review

Society of Automotive Engineers (SAE) International defines the degree of vehicle automation on six levels from level 0 (no automation) to level 5 (full automation) [*Automation Levels-SAE International* 2021]. An AMoD provides a MoD service consisting of a fleet of level 4 or 5 AVs, with level 4 AVs requiring geofencing and specific conditions for operation. In mainstream media, AMoD vehicles are alternatively referred to as *robotaxi* [*SF Officials Describe Street Chaos From Cruise, Waymo Robotaxis* 2023], *self-driving taxi* [WEDIA, 2021] or *driverless taxi* [HAWKINS, 2022].

With the widespread usage of MoD services and AVs on the horizon, the topic of AMoD gained significant momentum in the last two decades. An AMoD is typically seen as an amalgamation of multiple technologies that were in development for decades; a convergence of advancements in technologies like communication (e.g. smartphones, wireless data), navigation (e.g. real-time traffic updates, eco-routing), shared mobility, modern payment systems (e.g. mobile payments), electrification and vehicle automation [STOCKER and S. SHAHEEN, 2017]. Since most of the AMoD services only recently started their pilot phase, most of the research focused on simulations, surveys, or theoretical analyses of existing MoD and PT systems to study the potential impacts of AMoD services. The current section reviews the literature available for AMoD services. The section first discusses the history and current status of AMoD services (section 2.5.1), followed by a detailed analysis of AMoD service types (section 2.5.2) and modeling techniques (section 2.5.3) found in the literature. Since the dissertation focuses on improving AMoD fleet management and control, fleet management-related models are discussed separately in detail in section 2.5.4. The section mainly focuses on categorizing the services and techniques used in the AMoD literature. The overview of the literature reviewed using these categories is presented in Table A1 in Appendix A.

The section uses a similar taxonomy as adopted by [NARAYANAN, CHANIOTAKIS, et al., 2020; GOLPAYEGANI et al., 2022]. For the literature review of fleet management, some of the taxonomy is additionally taken from [M. F. HYLAND and MAHMASSANI, 2017; AGATZ et al., 2012]. Furthermore, the main focus is put on AMoD system modeling and operation, and thus, the long-term impacts of AMoD services on society in terms of factors like environment, travel behavior, land-use, are only discussed briefly wherever required. For a detailed discussion on these topics, the reader is referred to [NARAYANAN, CHANIOTAKIS, et al., 2020].

2.5.1. A Brief History and Current State of AMoD Services

The concept of AV has been around for a long time. In 1925, New York City residents were amazed by the sight of a vehicle driving through the streets of Manhattan without anyone at the steering wheel. A trailing vehicle controlled the vehicle via a transmitting antenna [STOCKER and S. SHAHEEN, 2017; *History of Self-Driving Cars Milestones* 2020]. In 1939, the concept received wider public exposure at General Motors' Futurama exhibit at New York World's Fair. There were further developments in AV technology in 1977 in Japan, which subsequently included Germany, Italy, the European Union,

---

[2]https://www.google.de/maps
[3]https://www.openstreetmap.org

and the U.S. [STOCKER and S. SHAHEEN, 2017]. However, the current interest in AVs concept could be traced back to the prized competitions by DARPA (from 2004 to 2007) [*The DARPA Grand Challenge: Ten Years Later* 2014]. In the 2004 challenge, the prize went unclaimed as none of the teams could finish the 142-mile course; nevertheless, during the second challenge in 2005, five vehicles out of 195 teams were able to complete a course of 132 miles. According to STOCKER and S. SHAHEEN [2017], Carnegie Mellon University, Environmental Research Institute of Michigan, and SRI International further strengthened the foundations of current AV developments. By 2013, several car manufacturers, including General Motors, Ford, Mercedes-Benz, and BMW, were working on their version of AVs. Google was already working secretly on AVs since 2009. Interestingly, many team members of Google also participated in the DARPA challenge of 2005 [CHAFKIN and BERGEN, 2017]. By 2016, over 30 companies worldwide were working on AV technology [STOCKER and S. SHAHEEN, 2017].

Besides the AV technology, the second most crucial element for the realization of AMoD services is the advancements in platforms for sharing vehicles. The earliest one among these is the car-sharing (CS) services. CS is a service model in shared mobility where the customer gets the benefits of a private car or light truck for a period of time without actual ownership and maintenance requirements of the vehicle. The vehicles are booked via a smartphone app [S. SHAHEEN, CHAN, et al., 2015]. Like AV, the traces of CS could be found much earlier than the modern developments. In 1948, the Sefage program was established in Zurich without much success due to easy access to private motorization [H. BECKER et al., 2017]. With developments in information and communication technologies (ICT) and mobile services, increase in fuel prices, and ever-increasing congestion on road networks, the CS systems started to get greater attention in the early 1990s and saw high growth in the early 2000s, especially from 2012 to 2014 when it witnessed a 55% percent growth in market share [S. SHAHEEN, COHEN, and JAFFEE, 2018].

To the best of the author's knowledge, the concept of combining AVs and CS[4] was first initialized in early 1990s in Europe, at the time named as Cybernetics Transportation System (CTS). It consisted of a network of purpose-specific AVs called cybercars that were limited in speed and area of operation. However, the researchers at the time remarked that the future of cybercars lies in its integration into regular cars [MICHEL PARENT and DE LA FORTELLE, 2005; M. PARENT and DAVIET, 1993]. The first instance of such a system was installed at Schiphol airport (Amsterdam) in 1997 for long-term parking to improve the service quality to airline passengers [JEANNETTE, 2020]. The system consisted of four electric *ParkShuttles* with three vehicles operating at a time while the fourth one remained charging and only called if there were additional transit requests. Later, in two decades, as the capability of AVs progressed to drive in the urban environment, so did the interest in AMoD services. As of January 2023, several big companies via their subsidiaries are competing to provide safe and sustainable AMoD services though they are still in the pilot phase and only operate in selected parts of the city. For example, Waymo LLC (a subsidiary of Alphabet Inc.; the parent company of Google) [ROTH, 2022] is currently offering AMoD services in Phoenix and San Francisco, USA. Similarly, Cruise LLC (a subsidiary of General Motors) [HAWKINS, 2021] operate in San Francisco and recently launched its services in Austin and Phoenix [BELLAN, 2022]. Zoox, Inc. (a subsidiary of Amazon) [MARASCO, 2023] is developing its own purpose-built robotaxis. Currently, their vehicles are being tested in San Francisco, Las Vegas, and Seattle.

Though the above pilot studies seem promising, there still appears to be a long way ahead till

---

[4]This combination describes AMoD services because, similar to CS services, the AVs are shared among customers, however, since the involved vehicles are AVs, the customers do not have to walk to the vehicle location; the AV can come to the customer. AMoD services can be equivalently described as a combination of AVs and current RH or RS services.
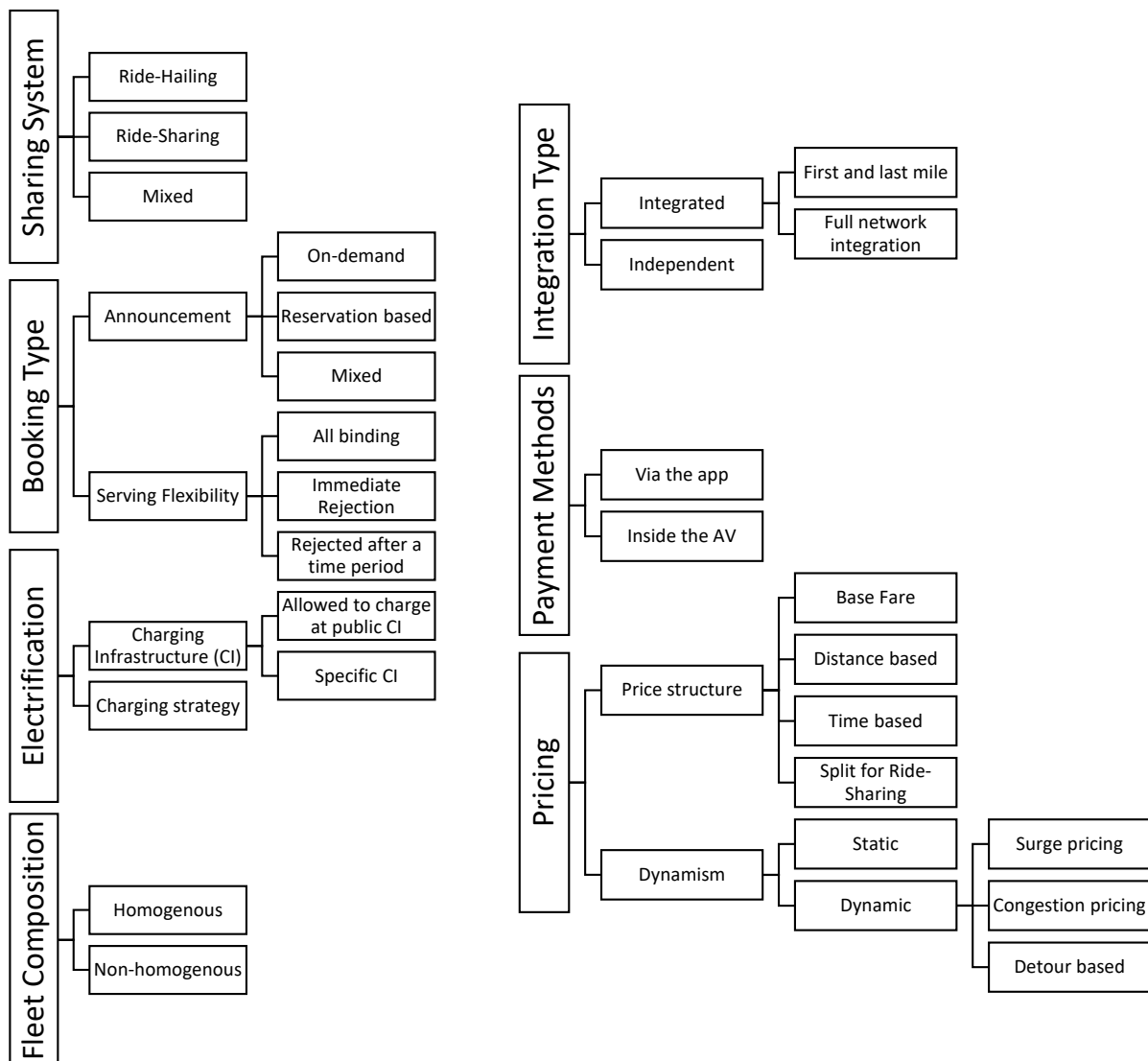
Figure 2.5.: AMoD service types.

wider acceptance of AMoD services due to safety requirements. For example, in one recent event on January 22, 2023, and AMoD vehicle from Cruise started to encroach into an area of emergency where firefighters were battling a major house fire. Two firefighters stood in front of the AV to prevent the vehicle from driving over hoses, and finally, they had to smash the front window of the AV to stop its movement. It is reported that there were 92 such unique incidents between May 29 and December 31 2022 [*SF Officials Describe Street Chaos From Cruise, Waymo Robotaxis* 2023].

2.5.2. AMoD Service Types

An AMoD service type defines the range of mobility features offered to the customers. These are used for advertising a specific AMoD service and building a brand name for the MSP. The more attractive these features and service quality are, the more popular the MSP gets, acquiring a higher market share. Since a large-scale AMoD is still non-existent, the literature gets inspiration from existing service models to study the impacts of potential AMoD service features. The exact features

offered by an AMoD could vary significantly due to its combining of years of developments in multiple domains, like electrification, payment schemes, shared mobility, and others. Over the years, each domain has developed its own service variants. The literature often brings a combination of these features together into a single system to study the impact of developed AMoD service type. Figure 2.5 summarizes the main categories for specifying an AMoD service type. The following explains these categories in more detail.

Type of Sharing System

An AMoD service could be classified into ride-hailing (RH), ride-sharing (RS) or a mixed system where the user can choose between an RH or RS offer. In the RH case, a single customer request is served at a time while in RS, a vehicle may pool multiple customer requests before dropping the ones already onboard. In terms of AMoD service, RH is also referred to as car-sharing (CS) for AVs [NARAYANAN, CHANIOTAKIS, et al., 2020]. RS will most likely be necessary for AMoD services to increase vehicle occupancy and reduce negative effects on the road networks and the environment [GREENBLATT and S. SHAHEEN, 2015]. Even though RS rides constitute a small share of current MoD rides [M. HYLAND and MAHMASSANI, 2020; HENAO and MARSHALL, 2019; S. SHAHEEN and COHEN, 2019; LAZARUS et al., 2021], the willingness of the customers to share rides in an AMoD system can only be known with certainty in future. However, similar to the current MoD system, the time-cost trade-offs would be most probably a significant factor for willingness to share rides in AMoD systems [M. HYLAND and MAHMASSANI, 2020; ALONSO-GONZÁLEZ et al., 2021; LAZARUS et al., 2021]. Thus, optimizing the operation of RS services for time-cost trade-off will still play a major role in AMoD services. Several researchers have studied and tried to improve the efficiency of AMoD services with RS option and varying vehicle capacities, for example, two [M. HYLAND and MAHMASSANI, 2020], four [ENGELHARDT, DANDL, BILALI, et al., 2019; JUNG, JAYAKRISHNAN, et al., 2016], six [ALAZZAWI et al., 2018; ZWICK and AXHAUSEN, 2020] and upto ten [ALONSO-MORA, SAMARANAYAKE, et al., 2017]. However, the operation of RS services remains a complex problem with significant potential for further improvements.

Type of AMoD booking

An AMoD booking type defines the nature of the ride request made by a customer and the response received from the AMoD operator. In a way, they define the type of service the MSP promises its customers, and if not fulfilled efficiently, can lead to a negative popularity of the MSP. NARAYANAN, CHANIOTAKIS, et al. [2020] distinguishes different types of booking based on booking time frame: on-demand (real-time booking of trips), reservation based, or mixed. Most of the AMoD literature only considered on-demand requests [ALONSO-MORA, SAMARANAYAKE, et al., 2017; ALONSO-MORA, WALLAR, et al., 2017; M. HYLAND and MAHMASSANI, 2018; ENGELHARDT, DANDL, BILALI, et al., 2019; FAGNANT and KOCKELMAN, 2018; FAGNANT, KOCKELMAN, and BANSAL, 2016] and only a limited number of studies considered a pure reservation based system [MA et al., 2017; LAMOTTE et al., 2017; LEVIN, 2017]. A reason for a higher focus on on-demand requests could be that the current MoD services mostly use on-demand requests only. Secondly, a pure reservation based AMoD system would be identical to a static Dial-a-Ride problem (DARP) if all customer requests are already known in advance. Interestingly, in recent years the interest in static and dynamic DARP has also increased significantly [HO et al., 2018]. Nevertheless, if a certain level of dynamism is maintained with reservation requests (i.e., the reservation can only be within a specific time horizon), the system could provide an upper bound to the purely on-demand requests based AMoD [MA et al., 2017], enabling better planning of routes and schedules, and lowering the

required fleet size [H. WANG et al., 2014]. Thus, recently, there has been a growing interest in a hybrid system where the user can choose between an immediate ride request or reserving a ride in the future [ABKARIAN et al., 2022; DUAN et al., 2020; ENGELHARDT, DANDL, and BOGENBERGER, 2022].

Besides the time horizon, a further distinction in the nature of booking can be made based on the flexibility of the ride request by the customer or the operator. Currently, in MoD systems, after assigning a vehicle to the customer, the drivers can accept or reject the ride request [*How Uber Works* 2023; *How to Give a Lyft Ride* 2023]. In contrast, this step will be mostly absent in AMoD services as there are no drivers unless the AMoD services incorporate private AVs and the AV owner is given a choice to accept or reject the ride request. Hence, accepting or rejecting a ride request will mainly depend on the central AMoD FC. Similarly, the customer may choose to accept or reject the provided offer. Regardless of what is implemented in actual AMoD, the studies in literature adopt one of these possibilities depending on the overall research objective. For example, some studies assume all customer requests to be binding and must be served regardless of waiting. This helps to compare different algorithms and evaluate the level of service offered by a given fleet size [M. HYLAND and MAHMASSANI, 2018; FAGNANT, KOCKELMAN, and BANSAL, 2016; PAVONE et al., 2012]. Similarly, some studies assumed that the customer leaves the system if not served within certain period [LOKHANDWALA and CAI, 2018; SPIESER et al., 2016]. Others assumed that the operator could immediately reject a customer if certain service quality criteria (mostly maximum customer waiting time to pickup) could not be met [ALONSO-MORA, SAMARANAYAKE, et al., 2017; ALONSO-MORA, WALLAR, et al., 2017]. Yet others also modeled the option of rejection by the customer [AL-KANJ et al., 2020; DANDL, BOGENBERGER, and MAHMASSANI, 2019].

Electrification, Fleet Composition and Environment

Electrification, fleet composition and a concern for the environment will play an essential role in the wide-spread acceptance of AMoD services by the general population and the policymakers [GREENBLATT and S. SHAHEEN, 2015; GREENBLATT and SAXENA, 2015; FULTON and OGDEN, 2021]. A general outcome usually found in literature is that emission reduction highly depends on efficient application of electric vehicles (EVs) in AMoD [NARAYANAN, CHANIOTAKIS, et al., 2020]. For example, FULTON and OGDEN [2021] explored different scenarios and concluded that electrification of the fleet is vital for decreasing overall energy consumption, without which, any efficiencies achieved via automation are compensated by the increased VKT caused by AMoD. JONES and LEIBOWICZ [2019] developed an optimization model for an energy system that integrates electricity and transport sectors. They note that AMoD fleet are more attractive for electrification due to their high utilization rate, making the vehicles with high capital costs and low variable costs more favorable. Furthermore, they conclude that the environmental and economic benefits are even more significant if charging of AVs is optimally scheduled.

In light of the above, several studies have considered AMoD system with electric AVs fleet, referred to as electric AMOD (eAMoD) in the dissertation. Two important aspects of the eAMoD services are (1) charging infrastructure (CI) and (2) the strategy used to monitor and schedule the charging of vehicles. The following discusses the charging infrastructure used in literature, while the charging strategies are discussed in later sections under AMoD fleet management.

To provide a consistent eAMoD service, the EV fleet would require sufficient CI. The most basic approach for the eAMoD studies for the CI is to use publicly available CI data for both the locations, number, and power charging units [DANDL and BOGENBERGER, 2019; Y. ZHANG et al., 2022; FEHN, NOACK, et al., 2019; L. LI, T. PANTELIDIS, et al., 2021]. In this regard, with the increasing electrification of POVs, the question arises if the AMoD should be allowed to charge at the public

charging stations, especially when the charging events of POVs are often combined with parking of POVs [Wolbertus et al., 2021] and have to compete with other vehicles to find empty charging spots with least additional VKT [A. A. Syed, Fischer, et al., 2022; Basmadjian et al., 2019]. Y. Zhang et al. [2022] compared the efficiency of a depot based eAMoD CI to the system where eAMoD fleet was allowed to charge at the public CI. The charging event of POV was derived from real data. The study indicates a decrease in empty VKT when public CI is used without additional investment costs for owning multiple depots for charging. However, in the absence of data or if CI is not yet available in the area of interest, models are used that determine the required infrastructure to sustain eAMoD operation [Chen et al., 2016; Vosooghi et al., 2020]. These studies generally also include sensitivity analysis for different vehicle ranges, charging unit power, and their impact on the performance of eAMoD system.

Besides the CI, the fleet composition, model type, battery size, and vehicle capacities also play an essential role in AMoD and eAMoD services. Greenblatt and Saxena [2015] estimated that, besides the electrification, by deploying appropriate vehicle types according to the passenger occupancy (i.e., small vehicles for requests with one or two passengers and larger for higher occupancy), the emissions could be reduced by a factor of two. If privately owned AVs are allowed in the AMoD system, that would already lead to a non-homogeneous fleet, similar to current MoD services. Atasoy et al. [2015] presented preliminary results of a similar concept in Tokyo; however, the fleet was homogeneous that adopted its capacity according to the chosen use-case by the passenger. One of these use cases included acting as a mini-bus with fixed pickup and drop-off locations. Additionally, in mixed fleets, a preference by customers for one vehicle type over another can also play a crucial role in the overall performance. Dandl and Bogenberger [2018] simulated two different vehicle models and showed that a customer preferring one over the other increases VKT. In summary, the topic of the mixed fleet in AMoD is complex as there are a significant number of possible choices. Further complexity in models is required to understand their impacts fully.

Integration Type

The integration type of the AMoD system plays a vital role in distinguishing the AMoD literature. Narayanan, Chaniotakis, et al. [2020] defines two major integration types based on the dependence of AMoD on other transport modes: (1) independent and (2) integrated with PT. An independent system exists parallel to existing transport modes and does not necessarily interact with or consider them in their operation. The majority of AMoD studies found in literature fall under this category, for example: [Pavone et al., 2012; Fagnant, Kockelman, and Bansal, 2016; Fagnant and Kockelman, 2018; Alonso-Mora, Samaranayake, et al., 2017; Alonso-Mora, Wallar, et al., 2017; Engelhardt, Dandl, Bilali, et al., 2019; Bischoff and Maciejewski, 2016]. However, several studies have shown that the introduction of current MoD services have shifted the traveling behavior of people away from the PT to MoD, causing an increased load on the road networks [Qiao and Gar-On Yeh, 2023; Kong et al., 2020]. Usually, the accessibility to the PT network plays a major role in people not choosing PT over POV, which is often referred to as the first and last mile problem [Murray, 2003]. Therefore, the studies with PT integration have attempted to complement the existing PT network by using AMoD for the first and last mile and potentially provide improved mobility options in areas of low PT coverage [Pinto et al., 2020; Moorthy et al., 2017; Shen et al., 2018; Wen et al., 2018]. Some works have tried to integrate AMoD and PT for the entire network and not just for first and last mile [Salazar et al., 2018]

In addition to the above, an AMoD could also be integrated into a larger platform known as mobility as a service (MaaS) apps, where the customer simultaneously has access to multiple mobility options, including bike-sharing, CS, AMoD, e-scooter and PT [Liimatainen and Mladenović,

2021; KAYIKCI and KABADURMUS, 2022]. However, a knowledge of modal split is usually required to evaluate the benefits of MaaS. A modal split deals with questions like how users select the transport mode, how it varies over time, and how it affects overall urban mobility. Therefore, there is a growing need for literature dealing with combined mode choice models of multiple transport modes [NARAYANAN and ANTONIOU, 2023].

Payment Methods and Pricing

Regarding payment methods, the AMoD system is expected to follow the same digital payment strategies as current MoD systems. Regardless of the booking platform used (MaaS or individual AMoD app), the users would pay via the same app as used for calling the rides. Due to growing concerns about data privacy and new technological developments in e-commerce, possibly the current payment methods would be replaced by more secure block-chain based methods, such as MOBI Wallet[5], or Mobility Coin[BLUM et al., 2022]. So far, the literature assumes that all the AMoD users will use the app accompanied by the payment method for calling a ride. However, the data has shown that despite the prevalence of MoD services, the traditional taxis with street hailing option are not disappearing and still have a significant market share [LYU et al., 2021; D. WANG et al., 2021]. In the long run, AMoD system may also incorporate gesture-based street hailing option using new technological developments. The biggest challenge in this regard would be to judge the intent of the customer gesture accurately. If such functionality is implemented in AMoD system, there could be a possibility to directly pay inside the vehicle using debit or credit cards without downloading an app.

The price structure plays a vital role in attracting more customers and keeping the whole AMoD system profitable and sustainable. The two main components of the price structure would be variable fare, paid per distance traveled or time spent inside the vehicle and the base fare. The base fare is usually required to make the AMoD system profitable by compensating for the empty driven trip to pick up the passenger [A. A. SYED, DANDL, KALTENHÄUSER, et al., 2021] or to discourage short trips that could be made simply by walking or biking [WILKES et al., 2021; WEN et al., 2018]. For the RS use-case, it is also possible to split the costs with other passengers or get discounts on agreeing to share rides [S. SHAHEEN, CHAN, et al., 2015; KUCHARSKI and CATS, 2020]. The exact values of these fares heavily depend on the cost of operating AMoD system, which is still a debated topic in the absence of an actual wide-scale for-profit AMoD service. The most commonly predicted values in the literature range from $0.19/km to $0.30/km [NARAYANAN, CHANIOTAKIS, et al., 2020]. A similar hardware cost estimate of $0.18/km (without service and maintenance costs) was mentioned by Waymo's former CEO in an interview [MORENO, 2021]. Some researchers suggest that these estimates are still optimistic and often overlook some important aspects; for example, LITMAN [2023] proposes that the misbehavior of some AMoD passengers cannot be eliminated entirely, which might require cleaning every 5-15 trips and occasional repairs. He proposes this may incur an additional $0.33-$2 per trip, excluding the value of time and driving costs to the cleaning station. BÖSCH et al. [2018] also conclude that the success of AMoD business model also requires finding solutions to the problem of customer misbehavior while onboard.

The price structure could also be dynamic with individual price components dependent on multiple variables — *flexible* or *dynamic* pricing [M. F. HYLAND and MAHMASSANI, 2017; NARAYANAN, CHANIOTAKIS, et al., 2020]. A detour-based price in RS system is an example of flexible pricing [KUCHARSKI and CATS, 2020]. Similar to current MoD fleet[6], charging a higher price at the time of high fleet utilization would increase profit and make sure that people who really need a ride and ready

---

[5]https://dlt.mobi/
[6]https://www.uber.com/de/en/drive/driver-app/how-surge-works/

to pay more, rather than waiting for prices to go down, are able to get a ride [AL-KANJ et al., 2020; TURAN et al., 2020]. This temporal increase of prices is also referred to as *surge pricing*[TURAN et al., 2020]. Similar to user-based relocations in CS systems [WEIKL and BOGENBERGER, 2013; WEIKL and BOGENBERGER, 2015], the prices in AMoD services can also vary spatially to incentivize people going to low-demand regions for better balancing of vehicle supply and demand [AL-KANJ et al., 2020]. Similarly, in order to avoid traffic congestion in certain areas, the regulators could impose congestion pricing (road pricing for using certain network links at certain hours of the day) which would also increase the prices paid by customers for some itineraries [SALAZAR et al., 2018; GURUMURTHY, KOCKELMAN, and SIMONI, 2019].

2.5.3. AMoD Modeling and Evaluation

Several methodologies are used in literature to study the impact of AMoD services. Some of the AMoD aspects could be studied in surveys, which helps to model those AMoD aspects in later studies, for example, factors affecting user acceptability [NORDHOFF et al., 2018; F. BECKER and AXHAUSEN, 2017], perceived benefits and drawbacks of AMoD [D. LI et al., 2022], understanding the mode choice preferences and reasons for pooling [LAZARUS et al., 2021]. Another approach is to use mathematical and analytical models to get a quick estimate of the AMoD systems. For example, BILALI, DANDL, et al. [2019] developed an analytical model to study the influence of reservation, detour, and maximum waiting time on the probability of finding shareable rides. Later on, they used macroscopic fundamental diagram (MFD) to further extend the model for enabling the ability to estimate traffic impacts when an AMoD service is introduced in a city [BILALI, FASTENRATH, et al., 2022]. Similarly, SANTI et al. [2014] used a shareability network to quantify the benefits of introducing RS in Manhattan. The shareability network creates a graph of trips that could be pooled and uses a maximum matching algorithm to study the percentage of shareable trips. While these analytical models provide a good opportunity for the MSPs to quickly quantify various aspects of an AMoD system, they often use simplistic assumptions to make the overall problem solvable within limited computation time. They also lack the flexibility to study the impacts of crucial elements like user interaction, vehicle charging, fleet assignments, etc. Therefore, as described in the following, an agent-based simulation is usually the method of choice for studying the impacts of AMoD services.

The AMoD system consists of a series of complex components that ultimately models the interaction between humans, AVs, and the environment (including the road network) [JING et al., 2020; NARAYANAN, CHANIOTAKIS, et al., 2020; M. F. HYLAND and MAHMASSANI, 2017]. In addition to each one of these actors having a complex system in itself, further complexity is brought by the stochastic and dynamic nature of the whole system; a decision made at any point in time by one of the actors can have a significant impact on the performance of the entire system. Therefore, the most common method found in literature is to use multi-agent models to study AMoD services [JING et al., 2020; NARAYANAN, CHANIOTAKIS, et al., 2020]. An agent is a closed system capable of taking autonomous actions in the surrounding environment to achieve desired goals [WOOLDRIDGE, 1997]. These agents interact with each other in an agent-based simulation. Realistic modeling of all the actors and their interaction simultaneously is generally difficult. Thus, the studies in literature usually model some of the components in greater detail than others depending on the overall objective of the study, which consequently also defines the assumptions made in the corresponding research.

This section presents the most important components of an AMoD system and the commonly used modeling techniques. The main focus of the section is on the agent-based AMoD simulation, and the analytical models are only briefly mentioned if necessary. The taxonomy adopted in the section is mainly inspired by [NARAYANAN, CHANIOTAKIS, et al., 2020] and [JING et al., 2020].

Customer, Vehicle and AMoD operator models

Multi-agent AMoD models consist of three main types of agents: customers, vehicles, and the AMoD operator. Usually, a large number of agents are computationally more expensive to simulate. In the simplest case, the customers are modeled as having some announcement or request time, at which point their presence is known to the system and must be served at some later point by AMoD provider [Pavone et al., 2012; M. Hyland and Mahmassani, 2018] or walk-away if they are not served till certain time limit [Spieser et al., 2016; Wen et al., 2018]. More complex customer models also include the ability to reject an offer and leave the system immediately if the offer does meet certain service quality criteria such as maximum detour time [Alonso-Mora, Samaranayake, et al., 2017; Alonso-Mora, Wallar, et al., 2017], probabilistic rejection after some waiting time threshold [Engelhardt, Dandl, Bilali, et al., 2019], or in case of RH a maximum waiting time [Dandl and Bogenberger, 2019; A. A. Syed, Kaltenhaeuser, et al., 2019]. It is also possible to add additional details to customer models that could impact the overall system, for example, the time spent for boarding and disembarking of the customers [Engelhardt, Dandl, Bilali, et al., 2019].

In AMoD simulations, in simplest terms, the vehicles are modeled as independent points traversing the edges of the city network [Erdmann et al., 2021; Alonso-Mora, Samaranayake, et al., 2017; Alonso-Mora, Wallar, et al., 2017]. The vehicles are supposed to be fully autonomous and completely follow all the instructions of the FC. Usually, a fixed number of vehicles with limited maximum capacity (according to the AMoD service being studied) are initialized at the beginning of the simulation, which remain the same throughout the simulation [Engelhardt and Bogenberger, 2021; Engelhardt, Dandl, Bilali, et al., 2019]. To determine the initial number of vehicles, some studies add new vehicles in warm-up simulation whenever the maximum waiting time threshold is not met [Fagnant, Kockelman, and Bansal, 2016]. In more complex studies, especially the ones using traffic simulators, vehicles do not simply traverse edges; instead, the car-following model is used that can replicate more realistic movements; the vehicles interact with one another, which allows taking into account the traffic congestion [Dandl, Bracher, et al., 2017]. Typically, this is the maximum level of detail found in AMoD studies for vehicle models. The detailed modeling of the internal functioning of AVs would only cost more computational effort without bringing additional insight into the overall AMoD system.

The AMoD operator model serve as the main component in the AMoD simulations since all the movements of the AVs are controlled by the operator. The main component of the operator model is the central FC that takes care of the whole fleet management, including assigning vehicles to customers. The internal functioning of the AMoD is usually modeled in detail along with different methods to improve the efficiency of the whole system. The important components and methods used for AMoD operator model will be discussed separately in later sections.

Demand Modeling

The demand modeling estimates customer demand in the study area, which generates the mode share for the AMoD services. As the AMoD are still in the pilot phase with limited service area, no direct AMoD operator data is available to model the demand. Therefore, many studies use trip data of existing services like taxis or MoD to model customer demand in the simulation. For example, using NYC taxi dataset [Alonso-Mora, Samaranayake, et al., 2017; Alonso-Mora, Wallar, et al., 2017; A. A. Syed, Kaltenhaeuser, et al., 2019; Erdmann et al., 2021; Dandl, M. Hyland, et al., 2019] or using Chicago taxi dataset [M. Hyland and Mahmassani, 2018; Dandl, M. Hyland, et al., 2020; Liu et al., 2020]. The taxi dataset usually only has

the trip's start time, not when the trip is requested. Thus, the studies directly use the trip start time as the time when customers order AMoD ride. Furthermore, these datasets usually use spatial and temporal aggregation due to data privacy, which requires disaggregating the data before they could be used in the studies. An additional advantage of using these datasets is that they allow the researchers to directly compare the efficiency of a centrally controlled AMoD system against a decentralized human driver-based taxi system. A major assumption for using the taxi data is that it represents the spatial and temporal demand for mobility, which is also assumed to represent the future AMoD demand — an assumption which may not be necessarily true as AMoD mode share is a complex and continuously evolving topic. As pointed out by HARDT and BOGENBERGER [2020], a major disadvantage of using these datasets is that they underestimate customer demand in an operation area. Usually, the datasets only contain information on trips where a vehicle successfully serves a vehicle; however, a significant portion of customer demand remains unreported as their ride requests could not be assigned due to a lack of vehicle supply. HARDT and BOGENBERGER [2020] also suggested an approach to model this additional demand for CS services using the expectation-maximization algorithm. Nonetheless, this additional demand is usually not taken into account in the AMoD literature; often, the AMoD works focus on comparing different fleet control strategies, for which using consistent customer data for all algorithms is more important than accurate modeling of future demand. Consequently, directly using these datasets without combining them with any demand modeling algorithm is widely common in AMoD literature.

While the taxi and MoD data provide a good source for potential AMoD demand, unfortunately, it is not available for most of the study areas. Therefore, the stated preference (SP) surveys are also commonly used in AMoD studies for demand and mode share modeling [MARTINEZ and VIEGAS, 2017; DIA and JAVANSHOUR, 2017; JAVANSHOUR et al., 2019]. SP surveys helps to quantify how people would react under a new situation by asking them questions related to new policies, product, or services [ALONSO-GONZÁLEZ et al., 2021; LAZARUS et al., 2021; D. LI et al., 2022; NORDHOFF et al., 2018]. Other data sources can also be used for this purpose, including smart transit cards data, land use data (locations and characteristics of residential, commercial, and school areas), mobile phone data, and others [NARAYANAN, CHANIOTAKIS, et al., 2020].

Using the above data sources (including taxi and MoD data), the AMoD demand could be either a constant factor of all trips [CHEN et al., 2016; FAGNANT and KOCKELMAN, 2014; ENGELHARDT, DANDL, BILALI, et al., 2019; A. A. SYED, KALTENHAEUSER, et al., 2019] or based on demand models [NARAYANAN and ANTONIOU, 2023]. Some simulators like SimMobility [AZEVEDO et al., 2016] and MatSim [SEBASTIAN HÖRL et al., 2018] have integrated mode choices directly into the AMoD simulation using multiple simulation layers. Some studies also consider dynamic adjustments to mode choices in the AMoD simulations, such as on a day-to-day basis [SEBASTIAN HÖRL et al., 2019] or on real-time information [WILKES et al., 2021; ATASOY et al., 2015].

Transport Network and Traffic Model

The *transport network* forms the main world or the environment where the customers and vehicles exist and interact with each other. The *traffic model* specifies the route flow and travel times between origins and destinations — the interaction with the transport network [NARAYANAN, CHANIOTAKIS, et al., 2020]. Both can have a significant impact on the accuracy of the simulation.

The AMoD literature mainly consists of three types of transport networks: 2D continuous plane, grid network, and node-link network. As the name suggests, in a 2D continuous plane, the transport network is defined on a continuous plan within a certain boundary. All of the positions within the boundaries are available for vehicle and customer positions. As for the traffic model, the vehicles can move on straight lines using Euclidean distances [BURNS et al., 2012], or on Manhattan geometry

for replicating the vehicles of an urban area [M. Hyland and Mahmassani, 2018; Dandl, M. Hyland, et al., 2019]. The vehicle speed is usually assumed to be constant [Dandl, M. Hyland, et al., 2019] or can vary for peak hours [Burns et al., 2012].

In contrast, the grid network discretizes the 2D space within a boundary. Thus, the network consists of a regular grid where grid points or centroids of each grid cell serve as nodes and their connection (represented by straight lines) with the neighboring grid points or centroids as edges of the network graph [Fagnant and Kockelman, 2014; Chen et al., 2016; W. Zhang, Guhathakurta, et al., 2015]. Unlike a 2D continuous plane, the customers or vehicles can only exist on a node or the edge of the network graph. Regarding the traffic model, the vehicle routes and distances can be calculated using the Manhattan metric. A Manhattan metric (or distance) between two points is calculated by the sum of absolute differences of their Cartesian coordinates, representing a movement over city blocks. Overall, the grid network represents a city on a grid with multiple areas representing different parts of the city, such as downtown, urban, or suburban areas. The network speed is predefined and can vary according to peak and off-peak hours or according to the area modeled [Fagnant and Kockelman, 2014; Chen et al., 2016].

In a node-link network, the nodes are not equally distanced over a grid; instead, they can be located at unequal distances. The node-link model can be further divided into coarse and road-level networks. Examples of a coarse network include studies where the transport network only consists of inter-connected stations (or hubs) [Pavone et al., 2012; Spieser et al., 2016; Liu et al., 2018] or inter-connected regional centers [Heilig et al., 2017]. In coarse networks, the traffic models are usually represented by straight lines between graph nodes. The travel distances or travel times are calculated using Euclidean distance [Pavone et al., 2012; Liu et al., 2018], however, more realistic travel times could also be incorporated using external sources like macro-simulation [Heilig et al., 2017].

The road-level transport network produces a more realistic simulation. They are based on high-fidelity maps, with each network edge having an actual road geometry, length, and travel time (or speed). In the raw map data, the travel time of the edges is typically based on free-flow speed. With the wide-spread availability of open-source projects like OpenStreetMaps[7], the usage of road-level transport networks became very common in literature, for example, [Alonso-Mora, Samaranayake, et al., 2017; Engelhardt, Dandl, Bilali, et al., 2019; *Urban Mobility System Upgrade* 2015; Boesch et al., 2016; Engelhardt, Dandl, A.-A. Syed, et al., 2022; Alonso-Mora, Wallar, et al., 2017; Dandl, Bogenberger, and Mahmassani, 2019]. For the traffic model, in the simplest case, the free-flow speed can be directly used [A. A. Syed, Kaltenhaeuser, et al., 2019]. While this approach can help to compare the efficiency of different fleet management methods, the replacement rate and the benefits of introducing AMoD services could be highly overestimated. Therefore, the studies use multiplication factors (either constant or time- and link-dependent) based on various sources to model congestion. Some of these sources include, for example, historic data [Jäger et al., 2018], an independent traffic simulation [Fagnant, Kockelman, and Bansal, 2016], occupancy per link [*Urban Mobility System Upgrade* 2015], Google Maps data [Bauer et al., 2018] or data provided by the concerned city authorities [Kondor et al., 2019]. The route between the source and destination is found by solving the shortest-path problem (section 2.4). Instead of dynamically solving the shortest-path problem inside the simulation, some studies preprocess the entire travel time and distance table (of either the fastest or the shortest route) for all combinations of nodes and reuse them throughout the simulation. This significantly reduces the simulation time [Engelhardt, Dandl, A.-A. Syed, et al., 2022; A. A. Syed, Dandl, and Bogenberger, 2021].

---

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

In most of the above studies with road-level transport networks, the vehicles are simulated as dynamic points traversing the network edges with the given speed; the vehicles are typically unaware of other vehicles traversing the same edge. However, the road-level networks can have even greater details when combined with microscopic traffic simulators, such as SimMobility [Azevedo et al., 2016; Basu et al., 2018], Aimsun [Dandl, Bracher, et al., 2017], Vissim [Alam and Habib, 2018] and others. The combination allows having multiple lanes on each link with a realistic simulation of individual vehicles and the overall traffic. This also helps to simulate effects like congestion with higher accuracy. The congestion effects can also be studied by combining AMoD simulations with mesoscopic traffic flow simulators like MATSim [Maciejewski and Bischoff, 2018]; MATSim uses a queue-based traffic flow model that is capable of running large-scale simulations.

Integrating other Transport Modes

Depending on the integration type of the AMoD system, the transport network must incorporate a model for other transport modes. With an increased number of transport modes, the number of possible routes between an origin-destination pair can increase drastically for each customer, making the overall problem significantly more difficult to solve. The problem becomes even more complex when additional AMoD fleet operations, such as repositioning, are also taken into account. Since integrating other transport modes is not the main topic of the dissertation, the section only briefly presents the methods found in the literature.

Among the transport modes, the most important is the PT network. The PT network is often line-based with fixed schedules, which makes it quite different from the transport network discussed in the previous section. The literature consists of multiple ways of combining the two networks. A common approach found in literature is to consider the PT network to function almost independently and to only interact with the rest of the AMoD network at fixed locations (mostly PT stops) with some predefined time for mode transfer. For example, Shen et al. [2018] study the integration of AMoD with the bus network for the first and last mile in Singapore. The buses are assumed to move with fixed speeds on corresponding bus lines and the AMoD vehicles can only pick up customers from bus stops. Salazar et al. [2018] study the integration of AMoD service with the subway network in Manhattan. They maintain separate graphs for walking, the AMoD service, and the subway network. Every node in the subway network is connected to the closest node in the walking graph. They assume a fixed period of one minute to transfer to the subway. A similar modeling approach is used by Salazar et al. [2018] and Zgraggen et al. [2019] for the subway system in Manhattan, where walking is additionally used as a transport mode. Their approach also includes repositioning of AMoD vehicles without a limitation of using AMoD service only for first and last miles.

In addition to the above, a more realistic integration of multiple transport modes could be achieved in combination with traffic simulators. Many microsimulators inherently support numerous transport modes, which could be used alongside AMoD simulations to study the impacts on combined system [Chouaki and Puchinger, 2021; Nguyen-Phuoc et al., 2023; Basu et al., 2018]. Since the PT network is simulated more realistically — especially the buses as they travel alongside other traffic vehicles —, this approach provides better integration of different transport modes.

The integration of multiple means of transport with AMoD is a complex topic with growing interest in various areas of research; for example, one study introduced the method for joint redesigning of the transit network and the determination of required AMoD fleet [Pinto et al., 2020]. There is also a requirement for further research on how to guarantee a higher level of service and on-time arrival of AMoD vehicle when multiple transport modes are combined.

(a) Overall AMoD system

(b) Fleet management tasks
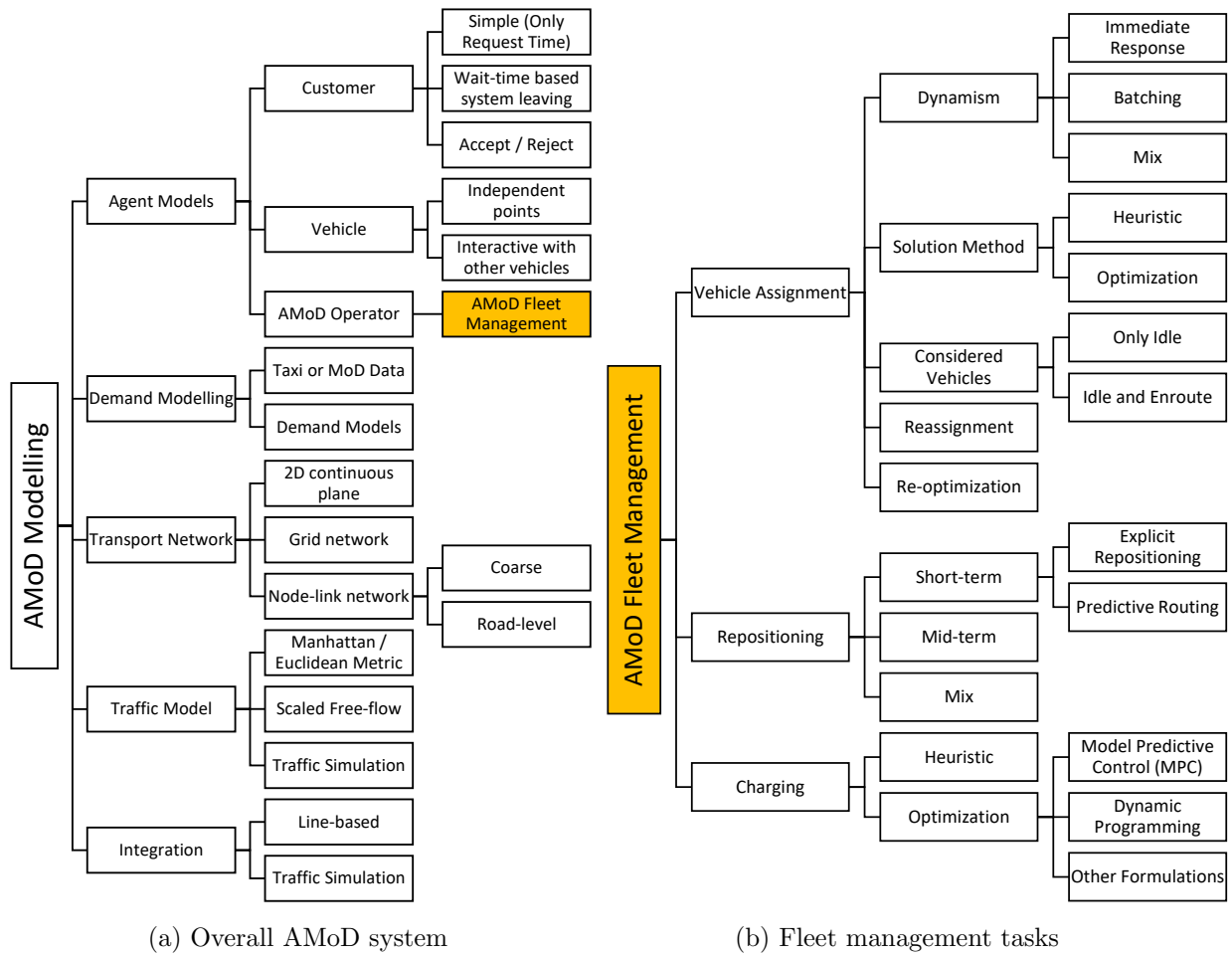
Figure 2.6.: Modeling of AMoD system.

2.5.4. AMoD Fleet Management

Unlike the current MoD services — where the human driver plays a crucial role and, in many cases, can make independent decisions —, the AMoD fleet is fully controlled by the MSP. The exact details of the fundamental driving tasks (speeding, braking, sensing, localization, and others) are still taken care of by the onboard software; however, the information of when and to where a vehicle should drive or which customer it should serve, must be communicated by the MSP. These functionalities come under the window of AMoD fleet management. The central fleet controller (FC) is generally responsible for fleet management at MSP. From the perspective of AMoD studies, the definition of AMoD fleet management also comes under AMoD modeling (section 2.5.3). Nevertheless, since the dissertation primarily concerns the AMoD fleet operation, the current section discusses it separately.

This section mainly reviews the three main AMoD fleet management tasks found in the literature: vehicle assignment, repositioning, and charging. Other control tasks could include dynamic pricing, dynamic fleet sizing or even controlling an AMoD based city logistic system [FEHN, ENGELHARDT, et al., 2021]. However, the literature is already limited on these topics from the perspective of FC and many of these studies have already been covered in the previous section.

In general, the overall fleet management of AMoD comes under the umbrella of VRPs (section 2.4). The history of VRPs goes back almost 70 years, during which hundreds of studies have tried to

formulate and solve different types of VRP. Looking at the long history of VRPs, M. F. Hyland and Mahmassani [2017] developed a taxonomy of the typical characteristics of the existing VRPs and the AMoD fleet management problem. Therefore, the following first discusses some of these characteristics before discussing the AMoD literature under each category of fleet management tasks.

VRP Characteristics and AMoD

The following only presents the major VRP characteristics relevant to AMoD fleet management tasks. The list is not exhaustive; the reader is referred to [M. F. Hyland and Mahmassani, 2017] for a complete list of characteristics. Some of these characteristics were also mentioned in previous sections. Nonetheless, the purpose of mentioning them here is to review how they are manifested in the mathematical formulations in FC.

- **Time-windows**: Depending on the AMoD booking type (section 2.5.2), FC may introduce certain time-window restrictions on fleet management tasks. If these restrictions are strict (*explicit* time-window) such that a customer *must* be picked up within that time-window, then the time-window is modeled as a hard constraint. This can happen if, for example, AMoD model allows reservations or has a certain upper limit on the maximum allowed pickup delay. Alternatively, if time-window is not strict (*implicit* time-window), then time-window constraints are added to the objective function, i.e., soft constraints. In this case, the time constraint is also called quality of service term.

- **Coupling and Precedence**: This occurs when a single request consists of multiple locations which must be served in a specific order; for example, a single shipment request might require collection from various points before delivery to the destination. The AMoD models generally consist of only two coupled points, i.e., pickup and drop-off points. In this case, it is called *pairing* instead of *coupling*. A higher number of coupled points could be imagined if the AMoD service allows picking up (dropping off) of the individual passengers in each AMoD request from (to) multiple locations. They are generally modeled as hard constraints. In VRP literature, these types of problems come under the category of pickup-and-delivery problem (PDP). If PDP involves the transportation of people, then it is generally referred to as Dial-a-Ride problem (DARP).

- **Vehicle Capacity**: A VRP may be constrained by the capacity in terms of weights, volume, number of places, etc. In AMoD, the capacity constraint occurs due to the maximum number of people allowed to sit in a vehicle simultaneously. Generally, the capacity concerns the RS case in AMoD, where the FC needs to keep track of current capacity while pooling multiple customers. It can also be relevant to RH case if the AMoD consists of a heterogeneous fleet with various sizes; the FC would need to make sure that a right size vehicle is assigned to the customer request. The capacity constraints also form a hard constraint in the problem. These problems come under the category of Capacitated Vehicle Routing problem (CVRP).

- **Information**: The information consists of any data required for solving the VRP. The type of available information plays a critical role in characterizing the problem. Some important aspects include:
  - *Quality of information*: The accuracy of the provided information can vary across problem formulations. Often part of the information is not known at the time of solving the problem. If part of the information is undetermined or described in statistical terms, then

the problem is *stochastic*; otherwise, if the information is known with certainty, then it is *deterministic* [Ho et al., 2018][Toth and Vigo, 2014, Chapter 11].

– *Evolution of information*: Many times, part or all relevant information is not known at the time of solving the problem, and the new information reveals over time. Thus, If the operator is allowed to modify or update the routing plan after the start of its execution, then the problem is called *dynamic*. In contrast, the problem is called *static* if all the related information is known beforehand. There can be multiple sources of dynamic information, for example, the addition of new requests into the system, availability of new vehicles, updated travel times, dynamic pricing, etc. A *dynamic* problem is also referred to as *online* or real-time problem.

The class of VRPs that has both of the above characteristics is termed as stochastic and dynamic vehicle routing problems (SDVRPs).

- **Objectives**: The objective function represents the goal that the FC tries to achieve. Some of the important aspects related to objective functions are:

  – *Single objective*: The objective function can consist of a single objective. The typical objectives include minimizing the number of vehicles, total travel cost, and total travel time or maximizing profit.

  – *Multiple objectives*: The problem can consist of a combination of objectives. The technique used for multiple objectives forms the basic classification, i.e., *weighted sum* approach, *hierarchical* optimization, or finding the *multi-objective Pareto front*.

  – *Selective visits*: If the fleet size is insufficient to serve all the requests, then only a subset of requests are served. Often problems with selective visits try to optimize multiple objectives either in the form of combined profit terms or via multi-objective optimization.

The fleet management in AMoD can have one or more of the above characteristics depending on the AMoD service model. However, some of the characteristics are inherent to AMoD fleet operations and are commonly found in the models in AMoD literature. First, from the perspective of the available information in AMoD system, the FC problems are inherently dynamic since none of the customer requests are known beforehand (unless AMoD model allows reservations). Additionally, the majority of AMoD literature includes some stochastic information into the FC, for example, via repositioning [Dandl, M. Hyland, et al., 2020; Pavone et al., 2012; Alonso-Mora, Samaranayake, et al., 2017], predictive charging using dynamic programming [Al-Kanj et al., 2020; Dandl, Fehn, et al., 2020], dynamic pricing [Turan et al., 2020; Al-Kanj et al., 2020] and others. Thus, the AMoD operational problem is a type of SDVRP problem. Second, from the perspective of coupling, it is a type of DARP. In DARPs, the users dial the operator for a ride from pickup to drop-off locations with the possibility of pooling the customers. More specifically, the AMoD could be termed as stochastic and dynamic DARP (SD-DARP). Ho et al. [2018] has reviewed the recent developments in DARPs which can also help to develop new ideas for the AMoD operations. Some of the research like [A. A. Syed, Kaltenhaeuser, et al., 2019; Santos and Xavier, 2015] already tried to utilize some of the DARP concepts to AMoD systems. Some noteworthy differences are still found in AMoD and DARP literature as described below.

The first difference is in the nature of the problem itself. Classically, the SD-DARP research deals with small-sized problems with large time-windows to accept or reject customers [Ho et al., 2018]. This exponentially increases the number of possible vehicle-customer combinations, making the problem significantly difficult to solve. Contrarily, the AMoD operation deals with a large number

of nodes in the city network characterized by a highly dynamic environment, where the FC has a very short time-window to accept or reject the customers. Furthermore, the customers in AMoD do not accept a very long waiting time [Lazarus et al., 2021]. Combining these factors with realistic travel times reduces the number of possible vehicle-customer combinations, making the overall problem easier to solve than traditional SD-DARP. The second significant difference is in the evaluation method used for solution approaches. Theoretically, any new dynamic method for SD-DARP (or general SDVRP) should measure its performance against the optimal solution of the static counterpart of the same problem instance, known as *competitiveness ratio* or *competitive analysis* [Ho et al., 2018; Bektas et al., 2014]. If the optimal solution is unknown, it is compared against the best-known solution. However, AMoD studies typically work with thousands of customers and vehicles simulated over a longer period, making it difficult to compare against the optimal solution of the static problem. They also use a variety of AMoD service models and data sources, making it difficult to reproduce the same simulation environment as other studies or compare against the best-known solutions.

Vehicle Assignment

The assignment of AMoD vehicles to customers is the fundamental task of an FC without which it cannot exist. Since the overall problem is an SDVRP, the first question raised is how to deal with the dynamic information. The FC operator has to make decisions based on the updated information. The typical approach found in AMoD literature is to accumulate customer requests for a fixed amount of time before solving a vehicle assignment problem — also referred to as *batching* of customer requests [Alonso-Mora, Samaranayake, et al., 2017]. The primary assumption for using batches is that by solving the overall problem in small optimization batches, the solution quality of the overall problem could be improved [Agatz et al., 2012]. In SDVRP and DARP literature, this method is also known as *rolling-horizon* approach [Bektas et al., 2014].

Alternatively, immediate responses are also possible where a vehicle is immediately assigned to the customer request or sent a preliminary confirmation without revealing the exact information of the assigned vehicle. However, the amount of literature available for immediate response is limited. Most of them use nearest-neighbor policy for this purpose [S. Wang et al., 2022; Erdmann et al., 2021; Liu et al., 2020]. Some studies also tried to combine immediate responses provided by the nearest-neighbor policy with batch optimization; however, the overall benefits of the combined approach over pure batch optimization have been limited [Erdmann et al., 2021; Erdmann et al., 2019]. A major limiting factor is that if the search space is pruned using a simple heuristic procedure (i.e., the customer is included or excluded from the optimization based on the immediate response by the heuristic), then the overall solution quality is decreased. On the bright side, this almost halves the response time to the customer due to immediate response.

In view of the above, the method of choice in the AMoD literature is to use various solution approaches along with batch optimization. The main question in this regard is which customer-vehicle pairs should be considered inside the batch optimization as many vehicles might already be serving customers when new customers are added to the system. M. Hyland and Mahmassani [2018] evaluated the impact of various possibilities (in total six policies) for the RH case in AMoD: two first come first serve (FCFS) based and four optimization-based policies. The latter involved a variation over if the customer-vehicle is *fixed* after first optimization or the customer could be *reassigned* to another vehicle (the vehicle enroute pickup would be reassigned to another customer). Additionally, the policy variation depended on if only the idle vehicles are included in the batch optimization or if vehicles enroute to drop-off are also considered. They concluded that reassignments and including enroute drop-off vehicles improve overall performance. It is also possible to use heuristics for batch

optimization, such as tabu search [ERDMANN et al., 2019], large neighborhood search [A. A. SYED, KALTENHAEUSER, et al., 2019], only consider vehicle within a certain radius for optimization [DANDL and BOGENBERGER, 2019] and others.

The problem complexity of AMoD service with RS is significantly higher than RH. Therefore, many researchers often use simpler insertion heuristics to build vehicle paths [HÖRL, 2017; FAGNANT and KOCKELMAN, 2018; GURUMURTHY and KOCKELMAN, 2018; SHEN et al., 2018]. ALONSO-MORA, SAMARANAYAKE, et al. [2017] presented the current state-of-the-art approach for large-scale ride-sharing scenario inside Manhattan using NYC taxi data. They first created a graph for the feasible pooling trips and then solved a set cover problem for assigning trips to vehicles. They assumed a hard-time constraint over maximum waiting time to create the feasibility graph inspired from [SANTI et al., 2014]. Others tried to use a similar method in other cities or improved the method further [ENGELHARDT, DANDL, BILALI, et al., 2019; ENGELHARDT, DANDL, and BOGENBERGER, 2020]. M. HYLAND and MAHMASSANI [2020] also presented an interesting approach where they showed that the simple assignment formulation could be used for RS problem if the vehicle is constrained to pooling only two requests at a time. Additionally, the method used can also change depending on the service model; for example, FIELBAUM et al. [2021] modified the method of ALONSO-MORA, SAMARANAYAKE, et al. [2017] to include a system where the RS customers are asked to walk to specific pickup locations.

Generally, for the SDVRPs problem, it is known that the incorporation of stochastic information into proactive approaches generally provides better performance than a myopic, reactionary approach [BEKTAS et al., 2014]. Such stochastic information can have multiple sources like stochastic travel times, blockage of some road links, future customers, and others. The AMoD literature generally uses customer-related information like arrival times, geographical locations, or overall demand patterns. Accordingly, several AMoD methods have tried to develop non-myopic vehicle assignment methods. ALONSO-MORA, WALLAR, et al. [2017] extended the method presented in [ALONSO-MORA, SAMARANAYAKE, et al., 2017] to manoeuvre AV paths to areas where they expect pooling more customers. DANDL, M. HYLAND, et al. [2019] tried to combine vehicle repositioning alongside vehicle assignments in batch optimization for RH. AL-KANJ et al. [2020] developed a dynamic programming based method that combines multiple FC tasks with vehicle assignment problem. Additionally, machine learning (ML) based approaches, though comparatively new to the field and limited to static problems, are still interesting for further research. Some studies like [NAZARI et al., 2018; HAMZEHI et al., 2019] tried to solve vehicle assignment problem using reinforcement learning (RL) while others [A. A. SYED, AKHNOUKH, et al., 2019; HOTTUNG and TIERNEY, 2020; SONG et al., 2020] improve DARP heuristics using ML for AMoD. Additionally, Y. WANG et al. [2019] take another interesting direction of using an adaptive batching framework using RL to decide when an optimization should be called. Most of these ML approaches still need to be either modified or tested in a realistic dynamic AMoD simulation.

Repositioning

As discussed at the end of the previous section, using stochastic information inside SDVRPs can significantly improve performance [BEKTAS et al., 2014]. The benefits of using future demand and anticipatory algorithms have been long known. POWELL et al. [1988] first utilized demand forecasts to develop an anticipatory approach for improving the performance of long-haul truckload trucking problem. They minimize the current and expected costs of future periods, showing a significant improvement compared to the pure reactionary method. Similarly, VAN HEMERT and LA POUTRÉ [2004] apply evolutionary algorithms in a rolling horizon fashion to solve the VRP with dynamic pickup loads. They model the advanced knowledge of the pickup loads in the form of regions. The

objective includes maximizing the number of transported loads while respecting the capacity and time constraints. They allow advance moves towards *fruitful regions*, even before the actual request is sent to the system. In general, the repositioning approaches have been important for system improvement. It reduces the preparatory or travel time required to reach the request point after it arrives in the system. They are essential for the ambulance services to cover large areas readily [Michel Gendreau et al., 2001; M. Gendreau et al., 2006].

In regards to the AMoD works, since it is computationally and statistically not feasible to deal with the exact geographical locations for customer forecasts, the general practice in the AMoD literature is to aggregate the customer locations into regions [Dandl, M. Hyland, et al., 2019; Dandl, M. Hyland, et al., 2020; Pavone et al., 2012]. Some researchers use self-defined zones for this purpose, while others use the zones defined by the authorities. The practice is similar to the one used for other VRP variants [Michel Gendreau et al., 2001; M. Gendreau et al., 2006; Bektas et al., 2014]. There can be multiple combinations of how this information is used in the AMoD operation cycle.

First, it can be used alongside the batch optimization — *short-term* utilization — either as a combined optimization problem [Dandl, M. Hyland, et al., 2019; Alonso-Mora, Wallar, et al., 2017] (as the previous section discussed) or right after solving the batch optimization [Alonso-Mora, Samaranayake, et al., 2017]. The anticipatory information can be either an actual forecast from past data [Dandl, M. Hyland, et al., 2019; Alonso-Mora, Wallar, et al., 2017] or derived information from not-served requests of the batch [Alonso-Mora, Samaranayake, et al., 2017]. The short-term utilization could be in the form of explicit repositioning to high-demand areas or maneuvering of the vehicles with onboard customers through demand-intensive regions (also called *predictive routing*) [Alonso-Mora, Wallar, et al., 2017]. Predictive routing is useful for RS case and is done by inserting hypothetical requests into the batch optimization.

A more common approach is to use repositioning completely separately from batch optimization. Usually, this is done periodically to rebalance different zones in the operation area. It is also sometimes referred to as *mid-term repositioning* [Dandl, M. Hyland, et al., 2019; Dandl, M. Hyland, et al., 2020]. While the above short-term utilization of statistical information can be considered a proactive approach to reduce the system imbalances in smaller steps, the mid-term utilization can be seen as a corrective measure for the accumulated imbalances. It can only be used with idle vehicles as modifying the paths of non-idle vehicles outside the batch optimization may unnecessarily disrupt the system. The problem aims at moving idle vehicles from areas where no demand is expected to areas where vehicles are likely to be needed. The idea is the same as relocation in CS systems [Weikl and Bogenberger, 2013; Weikl and Bogenberger, 2015], but does not require staff and is therefore much cheaper and frequent.

The *mid-term repositioning* strategy determines the number of idle vehicles that need to be repositioned from one region to another [Pavone et al., 2012; Dandl, M. Hyland, et al., 2020; Fagnant, Kockelman, and Bansal, 2016]. Usually, the objectives include either serving more customers or increasing the overall profit with the least possible increase in VKT. Nevertheless, most of the researchers apply their own procedures for its solution. For example, a common approach is to model the problem as fluid and try to redistribute the excess vehicles equally [Pavone et al., 2012; R. Zhang and Pavone, 2016; Ruch et al., 2018; S. Hörl et al., 2019]. Though the original approach [Pavone et al., 2012] used it with a station-based system, the later researchers generalized it using zone centroids instead of stations. Some studies utilize a *block-balance* defined by the difference between the share of available vehicles and expected demand for this purpose [Fagnant and Kockelman, 2014; Fagnant, Kockelman, and Bansal, 2016]. In this approach, the vehicles are repositioned to neighboring zones if the block balance is above a certain threshold.

Simpler heuristics like sending the vehicle to the closest zone or station are also found [SHEN et al., 2018; WINTER et al., 2021]. Predictive routing can also be used with mid-term repositioning. However, no work could be found where it is used. The main reason could be that maneuvering occupied vehicles is only beneficial in RS case and when done with higher frequency.

It should also be noted that the short-term and mid-term utilization of statistical information can also be combined. For example, in the short-term repositioning model of [DANDL, M. HYLAND, et al., 2019], they use a hard constraint on the maximum distance a vehicle can travel for repositioning. They notice that their short-term approach does not allow the vehicles to go to far-off regions, which causes many areas to remain imbalanced. Later, they combined the short- and mid-term repositioning in a dual horizon strategy to rectify the above problem [DANDL, M. HYLAND, et al., 2020]. Some researchers use more advanced models to look at the MoD imbalances from a holistic perspective. For example, ALBERT et al. [2019] use a stochastic differential model to define system imbalances as a difference of stochastic processes of request arrival and vehicle arrival. They use a distributed feedback control policy to average the system-wide imbalance.

It is also noteworthy that effective usage of the above strategies depends on accurately predicting customer demand. The above AMoD works utilized historic MSP data for predicting the customer demand; most of them use aggregated data of the not-yet-realized customer requests from the following time period — *perfect forecast*. However, the data of future requests are not available in actual operations. Discrete spatial information of users can help to improve these methods further. Works like [ALIPOUR et al., 2019] try to combine additional information like the location data of laptops and cellphones to predict exact user mobility. A growing concern for tracking user mobility without his consent and leakage of private data is leading to the development of *safer* alternatives. For example, [TONETTO et al., 2019] presents an initial step towards calculating crowd density estimations via the strength of the WiFi signal. However, these forecast methods should still be modified for AMoD usage.

Charging

Section 2.5.2 only described the approaches found in the literature for forming the charging infrastructure (CI) without describing the control strategies used by the eAMoD operator. These strategies help the FC to monitor the battery levels of the eAMoD fleet and determine when and where the AVs should be charged. Due to longer charge times, the overall eAMoD is more complex than a simple gas-based AMoD service. Consequently, the charging eAMoD fleet remains unavailable until a sufficient battery level is achieved, requiring the eAMoD operator to plan well for efficient performance. Due to the complexity of the problem, the vast majority of the available eAMoD literature used simpler heuristics for large-scale simulations. These heuristics often send vehicles to the nearest charging station available. For example, CHEN et al. [2016] uses FCFS principle to search for available AV within a travel time radius of 5 minutes (same as the simulation time step size) from the customer location. Before assigning the AV, if the AV range is not enough to serve the customer, the AV is sent to the closest charging station. The charging vehicle cannot be assigned to any customer. Later, they extended the approach to RS case [FARHAN and CHEN, 2018]. The downside of the method is that even a single customer request can initiate the charging process for a large number of vehicles in an instance, making the area suddenly under-supplied of vehicles. The extension of the method by [LOEB et al., 2018] can potentially solve this problem. First, finer monitoring of battery level is achieved by a smaller time step of 1 second. Second, any AV with a range below 5% or idle time of 30 minutes will be additionally sent for charging to the nearest station. Second, at the time of assigning customers, the policy makes sure that the AV additionally has enough range to reach the closest charging station after serving the request. The latter modification also allows the policy to

assign currently charging AVs to customers if all non-charging AVs are occupied. Some studies sent the AVs to charging if the charging level fell below a certain threshold [DANDL and BOGENBERGER, 2019; JUNG, CHOW, et al., 2014]. Some have also allowed charging enroute to rebalancing locations [L. LI, LIN, et al., 2019; T. P. PANTELIDIS et al., 2022]. Using heuristics, making extended plans for the charging stations is also possible. Y. ZHANG et al. [2022] combines the threshold-based charging method with a booking scheme for the CI. The eAMoD operator estimates the battery level after completing each vehicle's currently assigned route plan. If it is expected to fall below the threshold, eAMoD operator makes a booking at a charging station. Since each charging station also has other vehicles scheduled for charging, the eAMoD operator books the charging place with the earliest ending time for the full charge.

Even though the heuristic approaches provide a working solution for large-scale eAMoD services, their reactive nature lacks the foresight required for optimal charging performance. Therefore, an optimization-based charging procedure that incorporates the predictions of the future state is expected to perform significantly better. For example, a model predictive control (MPC) or dynamic programming (DP) based charging policies can be quite helpful. Some works like [R. ZHANG, ROSSI, et al., 2016; R. IACOBUCCI et al., 2019; AL-KANJ et al., 2020; DANDL, FEHN, et al., 2020] already applied these approaches to the problem; however, these methods still require further simplification to be scalable to a city-wide eAMoD service with hundreds of vehicles. Other optimization-based approaches are also possible. For example, L. LI, T. PANTELIDIS, et al. [2021] model the dispatching problem as a stochastic queuing problem and use minimum drift plus penalty (MDPP) approach to solve the customer assignment and the charging of vehicles in a single optimization problem. The method is also scalable to larger eAMoD services. A significant downside of the approach is that it breaks the first come, first serve principle; a customer coming into the system later might be served significantly earlier depending on the charging requirement of the overall trip. Nevertheless, the method shows the benefits of using an optimization-based charging strategy over heuristics.

In addition to the above, it is also possible to take into account the energy prices either in rule-based [FEHN, NOACK, et al., 2019] or optimization-based approach [DANDL, FEHN, et al., 2020]. The literature on the charging policies of eAMoD is continuously increasing. With further adoption of wide-scale eAMoD services, the corresponding research is expected to grow even faster with new topics.

2.6. Conclusion

The AMoD services are expected to produce the next paradigm shift in the transport sector when fully implemented. Looking at its potential, the AMoD research has been very active in the last decade. In addition to already existing service models, new service types and use-cases are continuously being added. In view of the impacts of current MoD services on the transport sector, on the one hand, the studies try to extract the maximum benefits from the new technology, on the other hand, the studies suggest being conscious of the negative impacts it can have if not commenced with foresight. Further studies are required on regulatory measures that could provide a smooth transition into AMoD future while mitigating potential negative impacts. More research is required on the impacts of AMoD as part of a multi-modal transport system. Additionally, due to environmental concerns, it is also essential to combine AMoD with electric vehicle technology. For eAMoD, in addition to the requirement of improving CI, further research is needed on the eAMoD charging strategies. It is also necessary to further investigate if CI can be shared between eAMoD fleet and POVs.

In addition to the charging strategies, there is also scope for further research on other fleet management topics. Even though there are a significant number of research available for RS use-cases, there

is a significant potential to further improve the RS performance, especially when combined with other transport modes and charging strategies. In terms of the studies that divide the operation area into a disjoint set of operational regions, there is a need for developing a consistent vehicle distribution model that does not depend on the assumption of independent regions. Such a distribution model can provide the opportunity to further improve fleet management tasks. The current dissertation is an attempt to fill this gap for RH use-case of AMoD system. The same vehicle distribution is then used in the dissertation to develop repositioning and assignment methods.

Chapter 3.

Basic Problem and Experimental Setup

This chapter presents the base AMoD service model and vehicle assignment problem focused on in the dissertation. First, section 3.1 introduces the basic AMoD service assumptions and the core optimization approach used for assigning vehicles to customers—vehicle control optimization (VCO). This fleet control optimization is used in the rest of the chapters for vehicle assignment. Later, section 3.2 presents the agent-based simulation framework used for AMoD performance evaluation. It also introduces the NYC taxi data set and the city network used for this purpose. Finally, section 3.3 introduces and evaluates a spatiotemporal scaling method that produces realistic travel times on the city network.

3.1. AMoD Service Type and Model

This section presents the formal description of the AMoD service type and associated assumptions used in the dissertation.

3.1.1. Studied AMoD service definition

The studied AMoD service has following characteristics:

- The assumed AMoD is strictly based on RH concept. This means that only a single customer request can be served by a vehicle at a time.

- Each simulation run of the AMoD scenario consists of a set R of customer requests.

- Each customer $r \in R$ is revealed at time t_r to the AMoD operator along with its pickup (p_r) and drop-off (d_r) locations. This is equivalent to r requesting a ride at time t_r via an app. The studied service model does not consider advance booking of a ride for r. Thus, the AMoD operator can only assign a vehicle to r after t_r.

- A customer pickup delay or waiting time is defined as the difference between the actual arrival time of the vehicle at p_r and t_r.

- A customer can only be picked up within the maximum waiting time, ΔT_{max}, starting from t_r. It is to be noted that the condition of ΔT_{max} is only checked at the time of assigning a vehicle to the customer. Since the dissertation uses dynamic travel times in the city network, it is possible that after the assignment of a vehicle, some requests are picked up with a delay of more than ΔT_{max} due to changing travel times. In such cases, the customer is assumed not to cancel the ride.

- Any customer that cannot be picked up before $\Delta T_{max} + t_r$ at the time of assigning vehicles will probably not be assigned a vehicle in the future. Thus, they are immediately rejected

and removed from the requests pool. This assumption is also theoretically proven in [DANDL, ENGELHARDT, et al., 2021] for RH services.

- The fleet V consists of homogeneous AVs that do not need recharging or refueling during the operation cycle. Thus, the dissertation does not consider the case of eAMoD.

- The assignment of vehicles to customers is controlled by a central FC.

- The size of V remains constant throughout a single simulation, i.e., dynamic fleet sizing is not used.

The dissertation uses the default value of 6 minutes for ΔT_{max} unless explicitly stated otherwise in the following chapters.

3.1.2. Service Quality Measurement

Since the assignment of vehicles to customers is an SDVRP, the overall performance or service quality of a fleet control algorithm can only be evaluated after all the customer requests have been simulated. The dissertation quantifies the service quality in terms of monetary profit, which is used to compare various algorithms at the end of the simulation.

As described in section 2.5.4, when the fleet size is limited and it is not possible to serve all the customers, then only a limited subset of the customers is served that maximizes the profit — *selective visits*. Several AMoD works adopted the same strategy to maximize the profit and maintain certain service quality in terms of the maximum allowed delay for picking up a customer (section 2.5.2).

As far as the monetary profit is concerned, it mainly consists of the overall revenue and variable and fixed costs. In reality, setting up an AMoD service would include several fixed and variable costs, such as maintenance costs for cleaning, refueling, or software infrastructure costs. However, for the purpose of the dissertation, i.e., comparing and improving the vehicle assignment strategies, many of these costs do not directly influence the used assignment method. Thus, the dissertation only considers the most relevant contributors to the monetary profit. First, each served customer r must pay a base fare of ζ and a variable fare of f^D per kilometer for the distance d_r^{pd} between pickup location p_r to drop off location d_r. The system considers a variable maintenance cost of c^D per kilometer and a fixed cost of c^F (total value for the whole simulation period, for example, for insurance and leasing) for each vehicle in the fleet.

With the above definitions, for a set of served customer requests $R^s \subseteq R$ and vehicle fleet V, the overall profit of the AMoD service is given as:

$$\sum_{r \in R^s} (\zeta + f^D \cdot d_r^{pd}) - \sum_{v \in V} (c^D \cdot d_v + c^F) \tag{3.1}$$

where d_v is the total distance traveled by vehicle v.

3.1.3. Vehicle Control Optimization

The dissertation batches or accumulates the new requests for a fixed period ΔT_{batch} before assigning them to vehicles. The FC solves an optimization problem to assign vehicles to each batch of requests, referred to as vehicle control optimization (VCO). Since the AMoD is a dynamic problem and the VCO only consists of a small portion of customers compared to all customer requests, the objective function inside the VCO can be different from the overall profit in Eq. 3.1 with additional terms for statistical information. While choosing an objective function for the VCO, the fundamental aim is
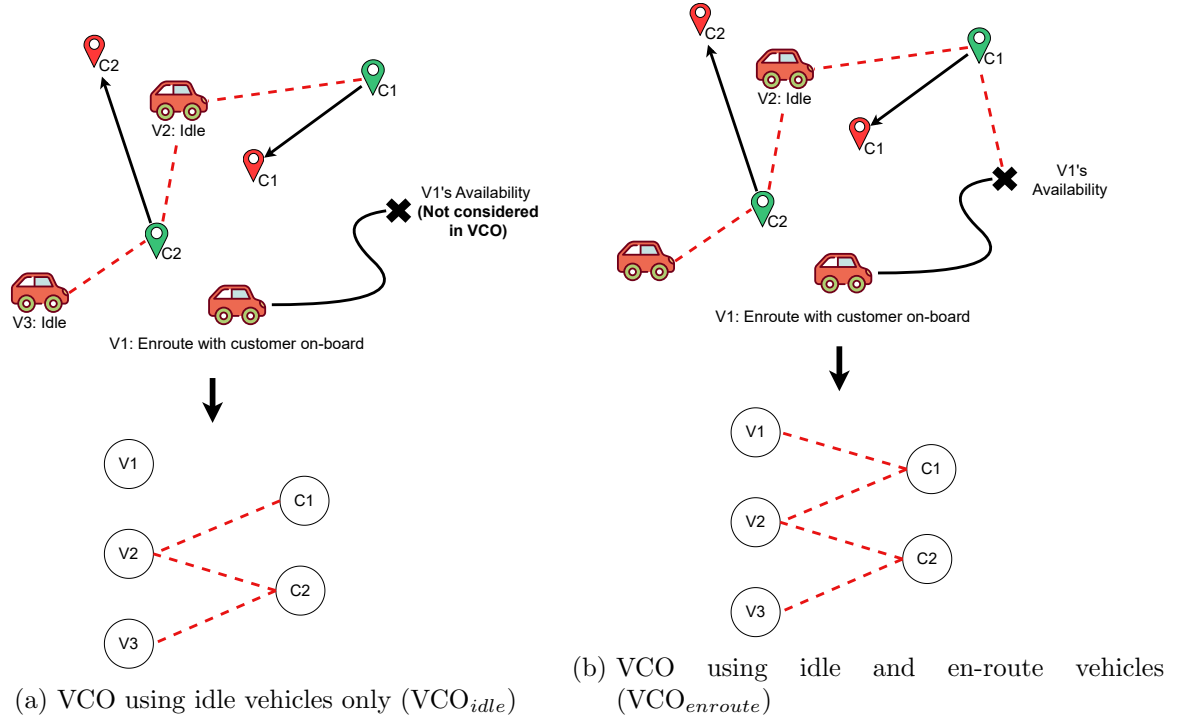
(a) VCO using idle vehicles only (VCO$_{idle}$)

(b) VCO using idle and en-route vehicles (VCO$_{enroute}$)

Figure 3.1.: The VCO types used in the dissertation. VCO$_{idle}$ only considers idle vehicles for customer pickups. In contrast, VCO$_{enroute}$ additionally considers the enroute vehicles based on their availability. The figure also shows the bipartite graph for each VCO type.

usually to improve the overall performance, which is measured according to a consistent metric — the monetary profit in the dissertation. However, for simplicity, the dissertation keeps the objective function in the VCO similar to the total profit in Eq. 3.1. Thus, the VCO for assigning vehicles to customer batches is described as follows.

Consider a set R_b^t representing a single VCO batch of customer requests at time t. Each customer $r \in R_b^t$ has an associated pickup location p_r, a drop-off location d_r, and the time t_r when the customer requested a ride. The time t_r also represents the earliest possible pickup time of r. Let the set $V^t \subseteq V$ represent a homogeneous vehicle fleet from which the VCO assigns vehicles to R_b^t. Each vehicle $v \in V^t$ can start a journey at time t_v from location o_v. Since at time t, some vehicles might still be moving, the literature consists of multiple approaches to determine which vehicles should be included in $v \in V^t$. For example, M. HYLAND and MAHMASSANI [2018] mention four variants to select V^t depending on if the vehicle enroute pickup or drop-off of customers is included and if the reassignment of customers to other vehicles is allowed. To focus mainly on utilizing spatiotemporal relations for performance improvement, the dissertation only uses two approaches to determine V^t, as shown in Figure 3.1. The first approach only uses idle vehicles at time t, referred to as VCO$_{idle}$. The second approach additionally includes currently enroute vehicles, referred to as VCO$_{enroute}$. In VCO$_{enroute}$, instead of the current time and position of the enroute vehicles, t_v and o_v of the enroute vehicles are set to the estimated time of arrival and location of the last point in the already assigned path, respectively. For the remaining vehicles without any assigned path, t_v and o_v are set to the current time t and vehicle locations, respectively. The dissertation also excludes reassigning already assigned customers to new vehicles for simplicity. Thus, a vehicle plan remains fixed after solving

VCO. For $VCO_{enroute}$, this means that the new vehicle plan could only be appended at the end of the already assigned vehicle plan .

To solve the assignment problem, the dissertation uses bipartite matching as described in section 2.4. First, a graph G with edges E for feasible vehicle-customer combinations is built. G consists of two distinct sets of nodes representing vehicles V^t and customer nodes R_b^t, respectively. There exists an edge $e_{vr} \in E$ between each vehicle node v and customer node r, if v can pick up r within a time-window of $[t_r, t_r + \Delta T_{max}]$, i.e. if the following condition is true:

$$\underbrace{t_v + \Delta T_{(o_v, p_r)} - t_r}_{=wait_{rv}} \leq \Delta T_{max} \tag{3.2}$$

where $\Delta T_{(o_v, p_r)}$ is the travelling time from o_v to p_r and $wait_{rv}$ represents the expected waiting time of r if picked up by v. Figure 3.1 shows an example of G for both VCO types, where the red dotted lines represent the edges of the associated G. In Figure 3.1, V3 does not have any edge with C1 because the condition in Eq. 3.2 is not fulfilled. Similarly, for $VCO_{enroute}$ in Figure 3.1a, even though the V1 is currently closer to the pickup location of C2, it has an edge with C1 because the condition in Eq. 3.2 is only applied from the estimated time and position where V1 will be available.

Let $R_G \subseteq R_b^t$ and $V_G \subseteq V^t$ represent the sets of customers and vehicles with at least a single edge in E, respectively. Then the VCO problem, with binary decision variable u_{vr} for each of the edge e_{vr} in E, is given as:

$$\max_{\boldsymbol{u}} \sum_{v \in V_G} \sum_{r \in R_G} (\zeta + (f^D - c^D)d_r^{pd} - c^D d_{vr}^{op})u_{vr} \tag{3.3a}$$

$$\text{s.t.} \quad \sum_{v \in V_G} u_{vr} \leq 1 \qquad \forall r \in R_G \tag{3.3b}$$

$$\sum_{r \in R_G} u_{vr} \leq 1 \qquad \forall v \in V_G \tag{3.3c}$$

where d_r^{pd} represent the distance from pickup to drop-off of r and d_{vr}^{op} the distance from availability point of v to the pickup location of r. The VCO objective in Eq. 3.3a is derived from the overall profit in Eq. 3.1 and maximizes the batch profit.

Some researches in literature also include a monetary cost for the value of time c^{VOT}, measured in monetary value per unit of time. This penalizes longer waiting times for pickups. However, the dissertation does not consider it in the VCO. First, c^{VOT} is an implicit cost whose actual monetary value cannot be clearly calculated. Its value is often adapted from surveys on how much people consider the worth of their time, which does not provide a concrete number and differ significantly from city to city [FREI et al., 2017]. Thus, the dissertation does not consider it in the overall MSP profit in Eq. 3.1. Even if the MSP profit does not use c^{VOT}, including in VCO might still be meaningful, especially for $VCO_{enroute}$. Without a penalty term for waiting time, the $VCO_{enroute}$ could, under certain circumstances, assign enroute vehicles to customers purely based on pickup distances from the availability point while an idle vehicle may have a slightly higher pickup distance but shorter customer waiting time. But as a penalty term in VCO, the c^{VOT} would require significant adjustments by the AMoD operator to meet the desired performance, in which case, the c^{VOT} would already loose its original meaning of being a monetary value for the time of customers wasted in waiting. It should be noted that even though the dissertation does not use a term for c^{VOT}, the waiting time of customers is still used as a hard constraint using ΔT_{max}.

The dissertation assumes that a request not matched with a vehicle in the current batch will most likely not be matched in future batches. For simplicity, all unassigned requests are immediately rejected and removed from future batches.

3.1.4. Key Performance Indicators

In addition to the key performance indicator (KPI) of monetary profit in Eq. 3.1, the dissertation uses several other metrics to compare the performances of different algorithms. These KPIs are also evaluated at the end of a simulation run.

1. **Customers served:** Generally, serving as many customers as possible is one of the main goals of vehicle assignment problems in AMoD services. The current MoD and Taxi services charge a significantly high base fare ζ than the variable fare f^D to make the overall business profitable, which may remain the same in AMoD services either for the same reason or to discourage short trips that could be covered via other means like walking [Wilkes et al., 2021]. This also makes the VCO prefer serving more customers. For comparing the performances of different VCO methods, it is also important to compare the total number of served customers in addition to the monetary profit. For this purpose, the dissertation uses the percentage of total customer requests served ($S_\%$) by the AMoD service as one of the KPIs.

$$S_\% = 100 \times \frac{|R^s|}{|R|} \tag{3.4}$$

2. **Waiting time:** The waiting time till pick up provides an important metric to measure the offered service quality. Even though the dissertation uses the hard constraint of maximum waiting time ΔT_{max}, it is still an important KPI to compare the performance. A lower waiting time would indicate that the AMoD system could position the available fleet closer to the potential customers and reduce their waiting times. Thus, the dissertation uses the mean waiting time of all the customers (W_{mean}) as a KPI for a single simulation run, given as:

$$W_{mean} = \frac{1}{|R^s|} \sum_{r \in R^s} (t_r^{pick} - t_r) \tag{3.5}$$

where t_r^{pick} is the actual pickup time of r by a vehicle.

3. **Empty distance:** For an overall efficient AMoD system, the fleet should be occupied by customers as much as possible. Any empty VKT traveled without an onboard customer causes an increased maintenance cost to the AMoD operator. In the studied AMoD system, there are two contributors to the empty distances traveled by the fleet: pickup distances and repositioning distances. Thus, the total empty distance is given as:

$$D_{total}^{empty} = \sum_{v \in V} (d_v^{repo} + d_v^{pick}) \tag{3.6}$$

where d_v^{pick} and d_v^{repo} are the sum of distances travelled by v to pick up assigned customers and reposition, respectively. D_{mean}^{empty} ($= D_{total}^{empty}/|V|$) is used for its mean value. Additionally, if D_{total} is the total VKT covered by the fleet, then the dissertation uses the notation $D_\%^{empty}$ for the percentage of empty VKT out of D_{total}.

4. **Pickup distance:** Similar to W_{mean}, a lower pickup distance indicates that the AMoD system can position the AMoD fleet closer to the pickup points of the customers. However, since the variable cost is measured per unit distance in Eq. 3.1, the pickup distances provide an important KPI for performance comparison. This is especially important for $VCO_{enroute}$, where the enroute vehicles are considered starting from the availability point and the waiting time depends on the finishing time of the last task in the vehicle's route. Thus, the dissertation uses the sum of the pickup distances of all requests from the respective availability point of assigned vehicles, given as:

$$D_{total}^{pick} = \sum_{r \in R^s} d_{vr}^{op} \tag{3.7}$$

where v corresponds to the vehicle that served r. The notation D_{mean}^{pick} is used for its mean value and $D_{\%}^{pick}$ as the percentage of D_{total}^{pick} out of D_{total}.

5. **Repositioning distance:** The repositioning VKT contributes to the empty distance the fleet covers to relocate to demand-intensive regions. This provides an important measure to compare the performances of various repositioning methods. The dissertation uses the notation D_{total}^{repo} for the total reposition distance of whole fleet and $D_{\%}^{repo}$ for the percentage of D_{total}^{repo} out of D_{total}.

6. **Fleet utilization:** For comparing the performance of the whole AMoD system, measuring how much the fleet capabilities were utilized is important. A lower fleet utilization would mean that the AMoD fleet was idle for a greater period. The dissertation defines fleet utilization in terms of the percentage of simulation time the fleet was busy. For a period of time Δt, it is defined as follows:

$$U_{\%} = 100 \times \frac{1}{|V|\Delta t} \sum_{v \in V} \Delta t_v^{busy} \tag{3.8}$$

where Δt_v^{busy} is the total amount of time v was busy (non-idle) during the period Δt. By default, the dissertation calculates the $U_{\%}$ for the whole simulation period (i.e., Δt is the difference between the simulation end and start times) unless the $U_{\%}$ is plotted on temporal axis in which case, the corresponding section mentions the Δt used. $U_{\%}$ can also be calculated for specific tasks like picking up customers or repositioning. The corresponding section mentions for which specific task it is calculated.

3.2. Experimental Setup

This section presents the experimental setup used throughout the dissertation. It first introduces the agent-based simulation framework used along with all necessary components. Then it discusses the method used to form the city network, followed by customer demand modeling using NYC taxi data set. Finally, it defines the division into operational regions used in the dissertation.

3.2.1. Agent-based Simulation Environment

The dissertation uses an agent-based simulation environment for testing the developed algorithms. The simulation consists of three main agents: customers, a fleet of AVs, customers, and AMoD operator. The simulation does not model competition among AMoD operators, and therefore, only single AMoD agent is used. The AV fleet is controlled by a FC located at AMoD operator. All simulations are done in the open-source AMoD simulator FleetPy[ENGELHARDT, DANDL, A.-A.
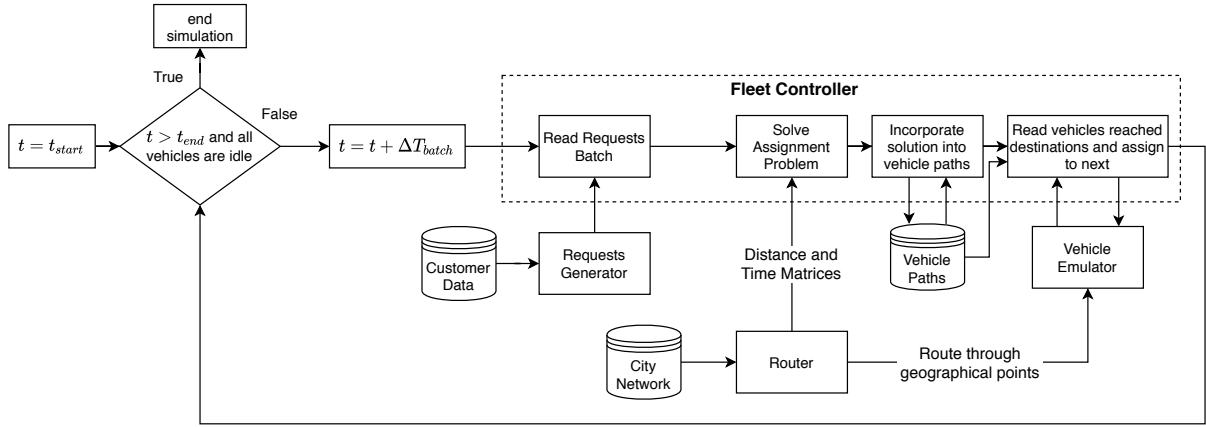
Figure 3.2.: Basic flow of agent-based synchronous MoD simulation

SYED, et al., 2022]. Figure 3.2 shows the basic flow of the simulation. The following summarizes some of its important aspects:

- **Time increments:** There could be two types of time increment schemes in a simulation environment for AMoD services: *synchronous* and *asynchronous*. In a synchronous scheme, all the simulation steps are done in series, including CPU-intensive operations like solving an optimization problem. In contrast, in an asynchronous scheme, the CPU-intensive tasks are kept in a separate CPU process or thread not to delay the operation of the rest of the simulation environment.

  For an AMoD simulation, the asynchronous scheme would mean that the VCO is solved in a separate CPU process while the rest of the simulation keeps running with smaller time increments [A. A. SYED, KALTENHAEUSER, et al., 2019; DANDL, BOGENBERGER, and MAHMASSANI, 2019]. This closely replicates the real AMoD scenario where the vehicles keep moving even when the VCO is being solved. In an asynchronous simulation, the FC puts a strict time limit for solving the VCO, which may not return an optimal solution within a limited time. This may also cause inconsistent evaluation of the control algorithms as the strict time limit on the VCO may change the assignment solution (if an optimal solution is not reached within the time limit), leading to very different results in each simulation run. Thus, this scheme only suits very fast online heuristics like the assignment of the nearest vehicle. It is also suitable when the simulation aims to evaluate the performances in an actual real-time environment whose simulated clock is intended to be close to the real clock. Such simulation studies would be interesting for actual MSPs, who want to see the performance improvement in the simulations closest to reality.

  In contrast to the above, in a synchronous scheme, the time increments in each iteration do not necessarily represent the actual clock. Some operations might take more time than the chosen time step size. This especially concerns the fleet movement and the solving of the VCO; the simulation time, and hence vehicle movements, is stopped while solving VCO. On first look, this might seem far from reality; however, if the simulation aims to evaluate some specific control algorithm for VCO, then it is essential to give more time to the VCO so that the full benefits of a control algorithm could be evaluated. The performance of a specific control algorithm can be improved separately.

  Additionally, if the batching of customer requests is used in a synchronous simulation and the

only dynamic event in the simulation is the arrival of new customers, then the time step size could equal the batching period ΔT_{batch}. This is a commonly used approach in AMoD literature [FAGNANT and KOCKELMAN, 2018; DANDL, M. HYLAND, et al., 2019; M. HYLAND and MAHMASSANI, 2018; FARHAN and CHEN, 2018]. Additionally, if the VCO takes less time than ΔT_{batch}, the synchronous environment can significantly speed up the overall simulation.

The possible differences between the overall results of synchronous and asynchronous frameworks could be minimized if the following are observed for each vehicle v at a simulation time t in a synchronous simulation:

- If v is idle or becomes idle after serving the last point in its assigned path and has no further route plan, then the vehicle availability time t_v is set to t.

- If v reached a point p_i at some time $t_v^{arrival}$ between $t - \Delta T_{batch}$ and still has remaining points to visit in its assigned route plan, then the serving of the next point p_{i+1} starts at time $t_v^{arrival}$ and not at t.

Since the dissertation focuses on comparing various control approaches, it uses the synchronous environment with the above mentioned corrections to get as close as possible to reality.

- **Requests generator:** In an actual AMoD service, the customers request rides dynamically via a smartphone app. The dynamic customers are simulated inside the simulation environment via a *requests generator*. Like many AMoD studies described in section 2.5.3, the requests generator uses historical trip data for modeling customer demand. It dynamically generates the customer requests for the current simulation time t. The historical data consists of customer pickup times and pickup and drop-off locations from past trips. The requests generator goes through this data and generates a batch R_b^t of customers for the period $[t - \Delta T_{batch}, t]$.

- **Vehicle Emulator:** The vehicle emulator (VE) consists of the simulated vehicles that are moved according to the time increments. The VE is not concerned with the long-term vehicle paths; it simply moves the vehicle to the following location provided by the FC. The long-term paths of each vehicle are kept in the FC. The *router* provides the geographical paths between two points on an actual city network to the VE.

- **Router:** The simulation environment uses an actual city network, i.e., a road-level transport network according to the taxonomy of section 2.6. However, access to the city network is only possible via a *router*, which forms the geographical information system (GIS) (section 2.4) for the simulation environment. The router provides the travel time and distance matrices between the geographical locations, which the simulation uses while solving the VCO. These matrices are calculated for the shortest distance paths between each location — not the quickest. This interface is used by each vehicle for traversing between two geographical locations.

  The router loads the distance matrix at the beginning and uses it throughout the simulation. To replicate accurate travel times, the dissertation scales the edges' travel times to match the historical travel times from past data, as described later in section 3.3. This is done by grouping the historical trips into periods of ΔT_{scale} (30 minutes in this work) and calculating the scaled time matrix for each period. This leads to multiple travel time matrices. The router loads the respective time matrix for each ΔT_{scale} period.

- **Programming language and City Network:** The whole simulation environment is programmed in Python 3.7. The drivable city (or street) network for the router is obtained from OpenStreetMaps (OSM) data via a Python library called OSMNx [BOEING, 2017]. The
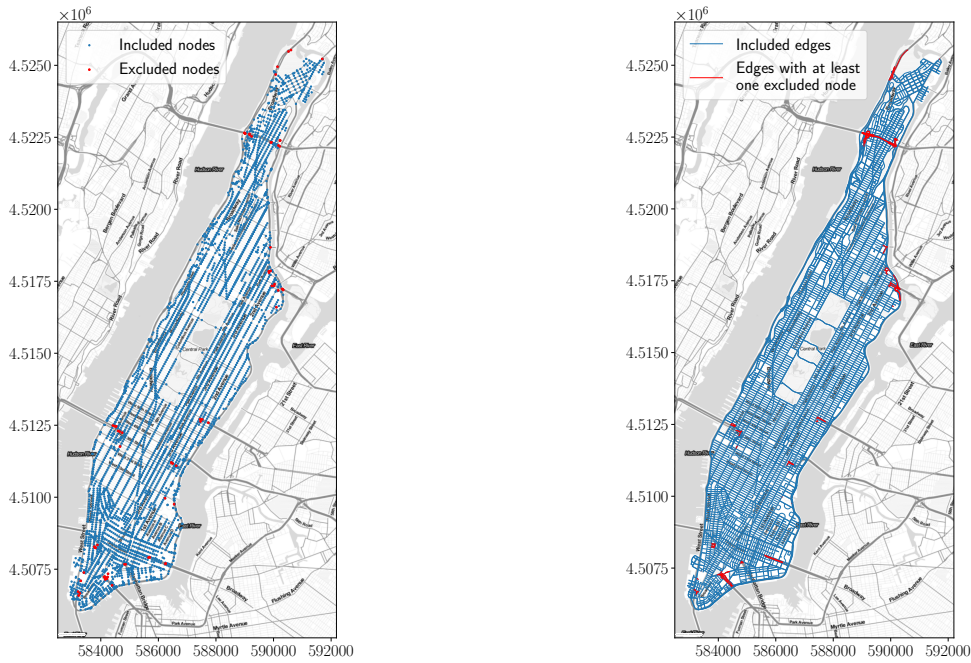
(a) 4508 total nodes (72 nodes excluded)   (b) 9737 total edges (128 edges excluded)

Figure 3.3.: The city network used for the Manhattan area. The excluded nodes cannot be used as a customer pickup or drop-off location or as an initial location of a vehicle.

library provides the city network in the form of a directed graph. To reduce the computational effort, the network is simplified using the built-in method of OSMNx. Since the vehicles can only traverse the edges between nodes, the dissertation makes sure that no vehicle ends up at a node without connecting edges. Thus, any node with an in- or out-degree of zero is excluded from the network. Similarly, some nodes may form an unreachable group from most other nodes even though they don't have non-zero in- and outdegrees. Thus, the dissertation also excludes any node that cannot reach (or cannot be reached from) more than 1000 nodes. The excluded nodes cannot be set as the initial location of a vehicle or a pickup or drop-off location of a customer.

With the above description, the dissertation first downloads the OSM map of the Manhattan area in NYC and preprocesses the network for the simulation. For the network edges with multiple lanes, the speed of the fastest lane is adopted. Additionally, some of the edges do not have speed information. The maximum speed of other edges with the same road type is used for these edges. If the network does not contain other edges with speed information and the same road type, then the speed of 25 mph (almost 42.3 km/h) is used. Figure 3.3 shows the city network used along with the excluded nodes. In total, the network consists of 4508 nodes, of which 72 nodes are excluded, and 9737 edges, out of which 128 are excluded as they are connected to an excluded node. Most of excluded nodes and edges are the ones that connected the roads going outside the Manhattan area.

(a) The number of trip origins
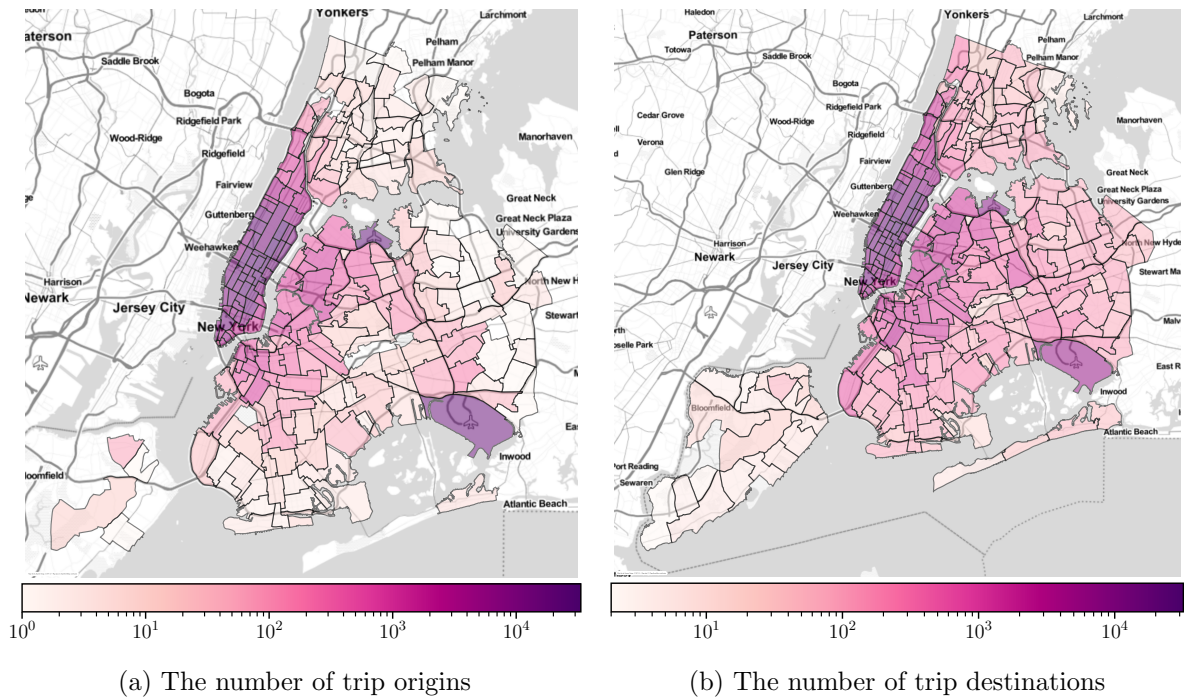
(b) The number of trip destinations

Figure 3.4.: The aggregated NYC yellow cab taxi data for 6th and 7th June, 2016. The data is aggregated over NYC taxi zones for display purposes only. The dissertation works with the direct coordinates provided in the data.

3.2.2. Demand Modeling

The dissertation aims to investigate various performance improvements for AMoD fleet management methods, which can also be achieved by generating random data. Random data is often used in operation research to evaluate the performances of multiple algorithms. However, since the dissertation focuses on performance improvement via the integration of statistical information, generating random data would have to follow patterns that at least match a realistic AMoD customer demand. Secondly, random data would not guarantee a similar performance improvement in an actual AMoD scenario. Therefore, a common practice found in AMoD literature is to use real user data from past trips of existing alternatives to AMoD services like taxis or MoD services for evaluations (section 2.5.3). The dissertation follows the same practice and uses the open-source NYC taxi data [*TLC Trip Record Data - TLC* 2023]. The data comes for two taxi types: yellow and green. The yellow cabs can pick up passengers from all the five boroughs[1] of NYC. In contrast, the green cabs are only allowed to pick up passengers from outer boroughs[2] (excluding the airports) and upper Manhattan. The dissertation uses the yellow cab taxi data, referred to as NYC data. For simplicity, the nearest nodes in the city network replace the trip origins and destinations. It should also be noted that even though the NYC dataset may underestimate the actual AMoD demand similar to other MoD datasets [HARDT and BOGENBERGER, 2020], the dissertation does not use any method to include those additional demands. Since the main aim of the dissertation is to develop efficient fleet control strategies, it is sufficient that the same data source is used for comparing different fleet control algorithms. This is also a common practice in AMoD literature (section 2.5.3).

---

[1]The Bronx, Brooklyn, Manhattan, Queens, and Staten Island
[2]The Bronx, Brooklyn, Queens, and Staten Island

(a) Hourly number of trips with line shadows showing daily variation (95% confidence interval).
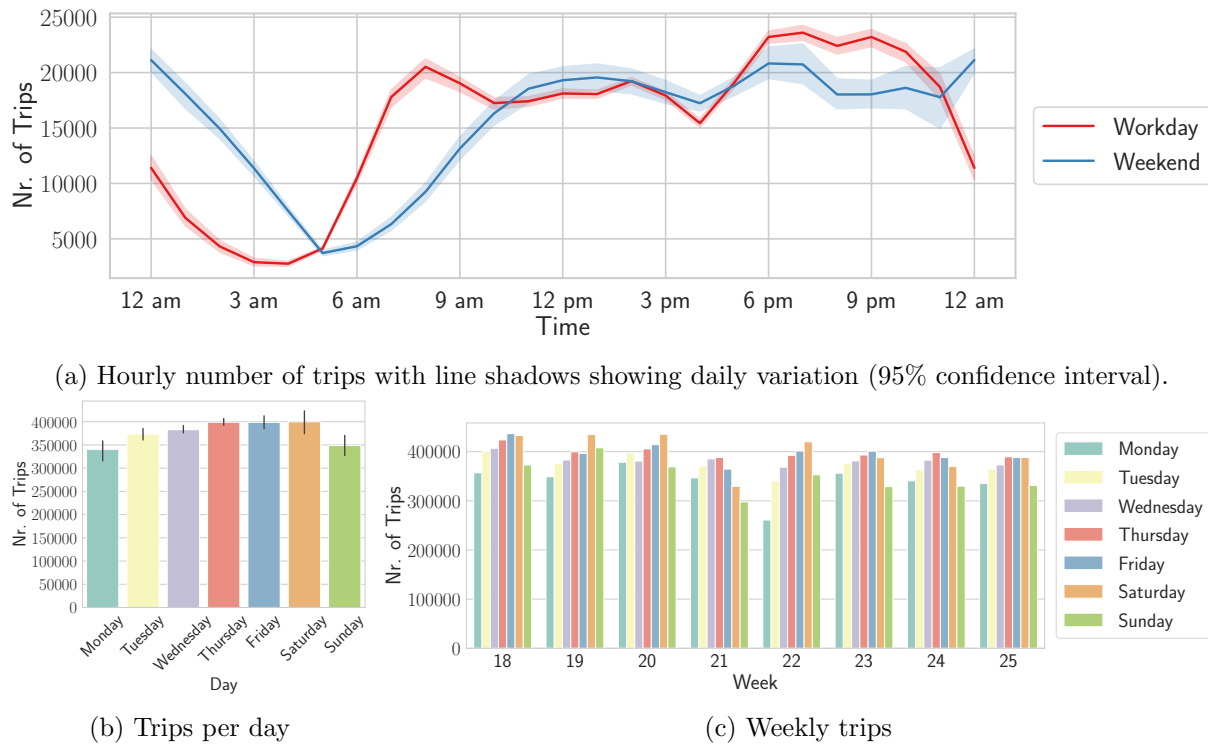

(b) Trips per day

(c) Weekly trips

Figure 3.5.: The NYC data from May and June, 2016.

The NYC data consists of billions of taxi trips which are updated regularly. The data till the middle of 2016 contained exact coordinates of trip origins and destinations. The later data sets aggregated the exact locations into areas of different sizes — called NYC taxi zones (Figure 3.4)— due to data privacy. The dissertation uses the NYC data from 2016 due to the availability of exact pickup and drop-off locations which benefits the evaluation of the statistical approaches used. The exact coordinates also allow replicating realistic network travel times as described in the next section. It should also be noted that even though the data used is almost seven years old, it does not affect the results as long as the same data is used for comparing all FC methods.

The data includes trip start and end times, origin and destination locations, and trip distances. The dissertation first removes the probable false trips that have a speed of less than 1 mph (almost 1.61 km/h) or more than 55 mph (nearly 88.5 km/h). Figure 3.5 shows the number of trips for May and June 2016 after removing false trips from NYC data. First, as shown in Figure 3.5a, the hourly pattern of workdays and weekends differs. The workdays are usually marked by two peaks: one in the morning when people go to work and one in the evening when people return from work. In contrast, the increase in the number of trips in the morning is rather smoother on the weekends. This corroborates the general behavior that people usually wake up late on the weekends and go out for leisure activities near midday. Additionally, the number of trips follows some pattern for each day of the week. On Mondays, the number of trips is usually smaller than other workdays, as shown in Figure 3.5b. However, there could be anomalies in this pattern due to various reasons. For example, as shown in Figure 3.5c, the pattern has an anomaly on Monday in the 22nd week due to a federal holiday on the 30 May.

Since the dissertation aims to test the studied methods in the most common cases, it avoids the days with anomalies. Thus, two days from week 23 are chosen; Monday and Tuesday (i.e.,
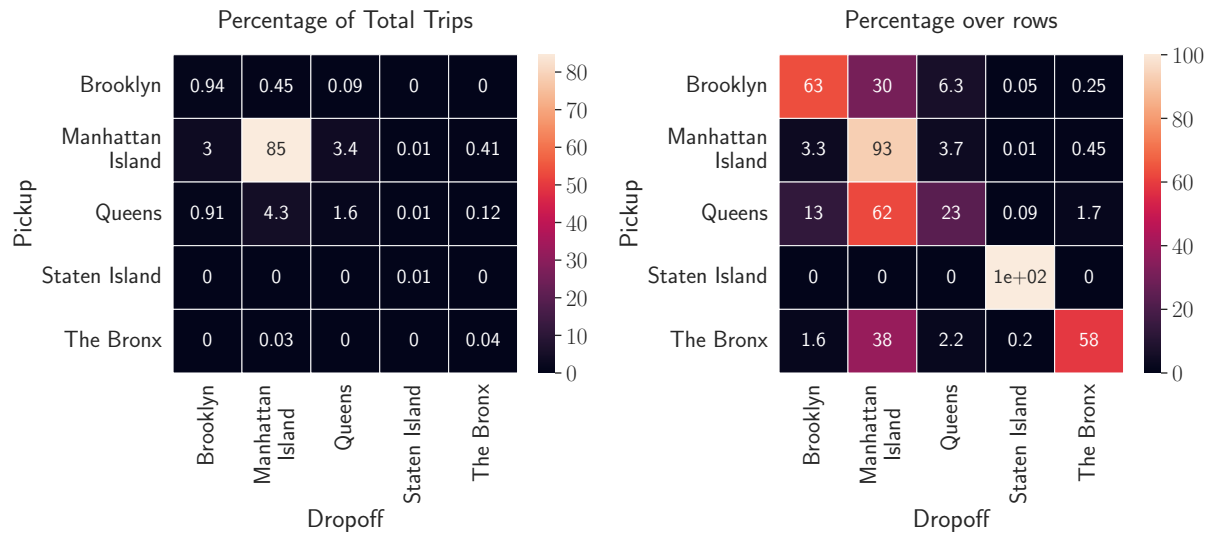
(a) Number of trip pickups and drop offs in NYC boroughs
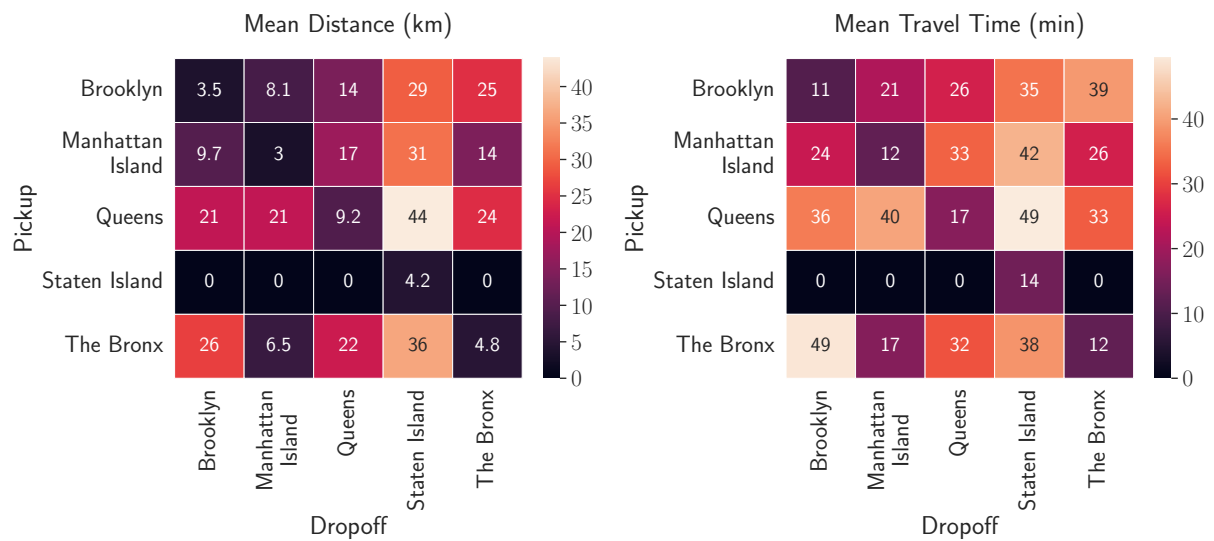
(b) Mean distance and travel times based on pickup and drop offs locations

Figure 3.6.: NYC data from 6 to 7th June, 2016.

6 and 7 June 2016). The specific choice of week 23 is rather arbitrary and any other week would be equally applicable. Additionally, the reason for choosing two days instead of just one day is to produce as realistic a simulation environment as possible; a single day of simulation may spend a significant amount of simulation time to make sure that vehicles are naturally distributed. This forms the warm-up phase of the simulation which may last a couple of hours. This is especially important since the pattern of trip origins and destinations differ significantly throughout the day and night. Longer simulation time guarantees that any improvement observed for an FC method is least affected by the initial positions of the fleet. On the other hand, a very long simulation time, for example, a week, would only increase simulation time without providing any significant improvement. Thus, the two days of simulation provide a good middle ground. The impacts of an FC method at rush hours or during the night could be observed on the second day when the simulation has already run for 24 hours. This makes sure that an FC would behave the same way if allowed to run for longer simulation hours.

In terms of regions, the NYC consists of five boroughs. The distribution of trip origins and destinations as well as the general features of trips vary among boroughs. Figure 3.6 describes the NYC data from 6 to 7 June 2016 based on boroughs. As shown in Figure 3.6a, almost 91.9% of the overall trips originate in Manhattan, and 92% (85% of the overall trips) of them have their destinations inside Manhattan as well. This serves as one of the main reasons why Manhattan was chosen as the study area in section 3.2.1.

The second highest trips, almost 6.78% originate in Queens, out of which almost 4.1% have destinations in Manhattan. Queens has two major airports of NYC: John F. Kennedy International Airport in the south and LaGuardia Airport in the north. This also explains why the majority of trip origins and destinations outside of Manhattan are concentrated at these two locations, as also illustrated in Figure 3.4. The trips starting from Manhattan also have the second highest destinations (3.7%) in Queens, as shown in Figure 3.6a.

Besides, the number of trips in each borough, the features of trips are also quite different among boroughs due to their sizes and inter-borough distances. As shown in Figure 3.6b, the mean trip distance and travel time for trips originating and ending in Manhattan are just 3 km and 13 minutes, respectively. While the trips that have either origin or destination in Manhattan (forming the second largest type of trips in the data set) have much higher trip distances and travel times: 17 km and 35 minutes for trips from Manhattan to Queens, and 21 km and 43 minutes for trips from Queens to Manhattan.

3.2.3. Division of the City Network into Regions

As mentioned in section 3.2.2, the NYC taxi data uses NYC taxi zones (Figure 3.4) to provide spatially aggregated information regarding trip origins and destinations. However, the NYC taxi zones would hinder tapping the full potential of spatial relations based methods studied in the dissertation. Thus, in addition to using the older NYC data that provides the exact coordinates, the dissertation divides the city network into regions using a regular grid with cell size Δs_{cell} instead of using NYC zones. The axis of the regular grid is not parallel to the used coordinate axes; rather, the *minimum rotated rectangle* function of the shapely library of Python is used to generate a rotated grid. This function returns the smallest possible rectangle that envelops the entire geometry without any restriction on being parallel to the coordinate axes. As Figure 3.7 shows, this ensures that each cell contains as much area as possible of the network with a regular grid.

The Python library can also mark the cells that do not contain any network node. These cells are excluded from the simulation.
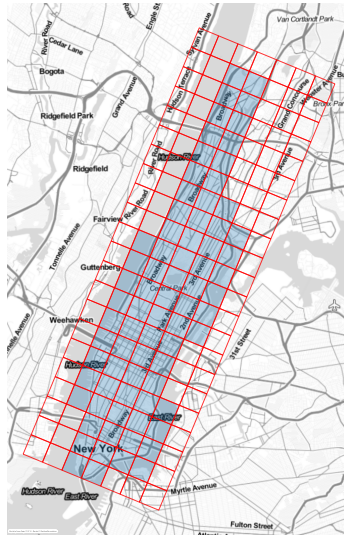
Figure 3.7.: An example of regular grid based regions with cell size Δs_{cell} of 1 km. The cells without color do not contain any node of the city network.

3.3. Scaling of Network Travel Times

Since the dissertation models the simulation vehicles as independent points traversing the network edges (refer to section 2.5.3), effects like congestion or realistic travel times are not inherently present in the vehicle model used. Therefore, the dissertation develops a method to scale individual network edges based on historical trip data to replicate a realistic simulation. The technique uses each historic trip to estimate the network state at different times of the day. It compares its efficiency with other commonly used scaling techniques found in AMoD literature.

The raw OSM network consists of free-flow travel time between edges of the network, which generally corresponds to the speed limits on road links. Many AMoD studies use some scaling technique to replicate a more realistic travel time. Section 2.5.3 mentioned some of the data sources used in literature for this purpose. Since the demand data set in the dissertation already contains exact coordinates, distance, and travel times, the dissertation uses this information to replicate the actual travel times in the data set. This section describes the methods that can be used for this purpose.

3.3.1. Mean Factor Method

This method divides historical trip data into periods of ΔT_{scale}. The division is made based on the recorded trip start time. Then, for a single temporal group, the mean factor method (MFM) scales all of the edges by a scaling factor [Dandl, M. Hyland, et al., 2020; A. A. Syed, Dandl, Kaltenhäuser, et al., 2021]. It first calculates the sum of the travel times of all historical trips, represented by t_{data}^{od}, within a single ΔT_{scale} period. Then the sum of the travel times for all trips is calculated using the free-flow speeds of the OSM network, represented by t_{osm}^{od}. The travel times of all the edges of the OSM are then multiplied by the mean travel time factor:

$$fact_{mean} = \frac{t_{data}^{od}}{t_{osm}^{od}} \tag{3.9}$$

MFM has the advantage that a single travel time matrix can be used for the entire simulation; the matrix is loaded at the beginning of the simulation and multiplied by the corresponding $fact_{mean}$ for the current simulation time. However, since it uses the total travel times for scaling, the operational regions with the highest number of trips can weigh significantly higher than those in other regions. Thus, it represents the actual travel times in high-demand areas much better than lower demand areas. This is especially problematic for the NYC data set, as many historical trips are concentrated in specific regions of the city.

3.3.2. Spatiotemporal Scaling Method

In contrast to MFM method, which only uses the temporal information to scale travel times, the spatiotemporal scaling method also uses spatial information for this purpose. Solving an optimization problem for each ΔT_{scale} allows a more accurate network link or edge-specific scaling. It is fundamentally based on the concept that each historic trip provides a snapshot of the network state at that particular time of the day. If multiple trips are grouped using temporal information, i.e., a period of ΔT_{scale}, they can potentially provide numerous observations of the overlapping edges on their paths from origin to destination, as shown in Figure 3.8. However, the method makes certain assumptions regarding the data set, which may limit its applicability, as described below.

First, it assumes that the data set also provides the exact paths taken by the historic trip. However, most of the data sets do not contain such information. The section solves this problem by assuming that the historic trip took the shortest-distance path between origin and destination. This can be easily calculated using the OSM network used in the dissertation. The assumption of the shortest-distance route could be quite realistic since many drivers in the trip data sets are usually experienced drivers and are well aware of the shortest paths. Many of them might have unintentionally taken the shortest-distance route, which would make the developed paths close to reality.

Second, since the routes are manually generated using shortest-distance routes of OSM network, it is not known when each edge might have been crossed within the period ΔT_{scale}. The method assumes that all trips within a group (formed using ΔT_{scale}) crossed the overlapping network edges within that specific ΔT_{scale} period. This is a strong assumption and may not be accurate for each group. Thus, a single network edge can have multiple values with significantly varying travel times. The errors caused by this assumption could be reduced using smaller ΔT_{scale}, as the network state will likely remain similar for shorter periods. Nevertheless, using a very small ΔT_{scale} would increase errors due to fewer observations for each edge.

Due to the above limitations, it becomes necessary to solve an optimization problem for finding scaling factors for the network edges that best fit the observation of network edges. The main goal is to scale the individual network edges such that the travel time difference between recorded travel times and the travel times obtained from the scaled network is minimized for most trips. The section develops two variants of the scaling method based on the objective function used for optimization.

The first method is referred to as the squared scaling method (SSM). It divides the whole operational area into regions Z using a regular grid of cell size Δs_{cell} (section 3.2.3). Let E_{net} represent the edges of the network used for AMoD simulation, t_e^{flow} represent the free-flow travel time of an edge $e \in E_{net}$ and $E_z \subseteq E$ represent the edges that are entirely within region $z \in Z$. Some of the edges may be located in multiple regions, denoted by $E_m \subseteq E_{net}$. For each of these edges $e_{\hat{z}} \in E_m$, the set $\hat{z} \subseteq Z$ represents the regions the edge belongs to. Afterward, it calculates the shortest paths for each of the trips within the period ΔT_{scale}. Let the set C_{data} represent all the trips within the period ΔT_{scale}, then the shortest path of a trip $c \in C_{data}$ consists of a sequence of network edges, represented by $E_c \subseteq E_{net}$. The main purpose of the method is to scale the free-flow travel time
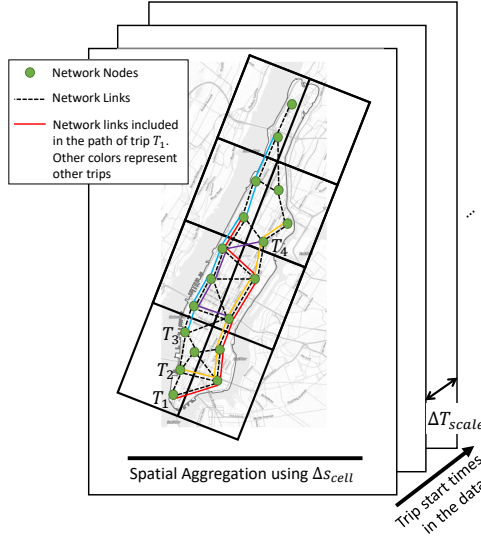
Figure 3.8.: An example of spatiotemporal grouping of customer data for scaling the city network. For simpler visualization, the directional edges between the nodes are reduced to a single edge.

of each edge in E_c for all $c \in C_{data}$ such that the travel times comes as close as possible to the historical travel time t_{od}^c for $c \in C_{data}$. If two or more trips in $c \in C_{data}$ have the same origins and destinations, then the mean of the travel times of these trips are used.

Let $E_u \subseteq E_{net}$ represent the union of all edges in the shortest paths of C_{data}, i.e., the union of E_c for all $c \in C_{data}$. Then, as the first step, the method only scales the edges in E_u by solving the following optimization problem.

$$\min_{\boldsymbol{x}} \sum_{c \in C_{data}} \left(t_{od}^c - \underbrace{\sum_{e \in E_c} x_e t_e^{flow}}_{\text{scaled travel time}} \right)^2 \tag{3.10a}$$

$$\text{s.t.} \quad x_e t_e^{flow} \leq t_e^{max} \qquad\qquad \forall e \in E_u \tag{3.10b}$$

$$x_e \geq 1 \qquad\qquad \forall e \in E_u \tag{3.10c}$$

$$\boldsymbol{x} \in \mathbb{R}^{|E_u|}$$

where $\boldsymbol{x}$ is a vector of linear variables used for scaling each edge within E_u and t_e^{max} is the maximum allowed travel time on the edge. t_e^{max} is calculated using the edge length (d_e) and a parameter for the minimum travel speed allowed on the edge (S_e^{min}), i.e., $t_e^{max} = d_e / S_e^{min}$.

After calculating the scaling factors for the edges in E_u, the squared scaling method (SSM) scales the rest of the edges in E_{net}. These additional edges are divided into two categories. First, the ones belonging to a single region (i.e., belonging to E_z) are scaled using the mean of the scaling factors already calculated in the first step for that region, i.e., they are scaled by the mean of all $x_e \in E_u \cap E_z$. Second, the edges that fall within multiple regions are scaled using the mean of the scaling factors for all edges within those zones, i.e., for an edge $e_{\hat{z}} \in E_m$ it is scaled by the mean of all $x_e \in \cup_{z \in \hat{z}} (E_u \cap E_z)$. The main assumption is that the scaling factor should be on a similar scale in the immediate surrounding of the scaled edges E_u. If any edge is still not scaled using these two approaches, they are scaled using the mean of $\boldsymbol{x}$.

The second spatiotemporal scaling method uses the same approach as SSM; however, instead of using squared error for the objective function, it uses absolute error. Thus, the method is referred to as absolute scaling method (ASM) and given as:

$$\min_{\boldsymbol{x}} \sum_{c \in C_{data}} \left| t_{od}^c - \sum_{e \in E_c} x_e t_e^{flow} \right| \tag{3.11a}$$

$$\text{s.t.} \quad x_e t_e^{flow} \leq t_e^{max} \qquad \qquad \forall e \in E_u \tag{3.11b}$$

$$x_e \geq 1 \qquad \qquad \forall e \in E_u \tag{3.11c}$$

$$\boldsymbol{x} \in \mathbb{R}^{|E_u|}$$

To solve the problem in optimization software, the following reformulates the problem to remove the absolute function.

$$\min_{\boldsymbol{x}} \sum_{c \in C_{data}} u_c + v_c \tag{3.12a}$$

$$\text{s.t.} \quad x_e t_e^{flow} \leq t_e^{max} \qquad \qquad \forall e \in E_u \tag{3.12b}$$

$$x_e \geq 1 \qquad \qquad \forall e \in E_u \tag{3.12c}$$

$$u_c - v_c = t_c^{od} - \sum_{e \in E_c} x_e t_e^{flow} \qquad \qquad \forall c \in C_{data} \tag{3.12d}$$

$$\boldsymbol{x} \in \mathbb{R}^{|E_u|}$$

$$\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}_+^{|C_{data}|}$$

where the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ represent the positive and negative error between the historical and scaled OSM travel times, respectively.

3.3.3. Analysis of Data used for Scaling Travel Times

The NYC data does not include the actual routes taken by trips. Therefore, both SSM and ASM estimate the original routes (required for scaling individual network edges) using the shortest distance paths of the OSM network. To have a rough estimate of the accuracy of the calculated routes, Figure 3.9 shows the difference between the recorded travel distances obtained from the data and the distances of the shortest paths calculated using OSM network, referred to as *offset distance*. The shortest path distances are calculated using the Universal Transverse Mercator (UTM) projection of the network. It is assumed that if the original trip took a similar route as the shortest path, then the offset distance must be small. However, as mentioned in section 3.2.1, the network is simplified to reduce its size, and the distances are measured using the nearest network nodes to origins and destinations — not exact trip coordinates which could also be in between two network edges. This can also contribute to an increased offset distance.

With the above considerations, 73% of trips have an offset distance of less than 250 m, and almost 92.4% of the trips have a difference of less than 1 km, which confirms our initial assumption that the historic trips routes can be estimated using shortest-distance routes. The proportion of trips having higher offset distances is significantly lower. The trips with higher offset distances could have either taken a different route or visited multiple stops at different locations on their way to the destination. The negative offset distance could mean that the actual geographical locations for origins and destinations could be somewhere on the network edge with a smaller travel distance
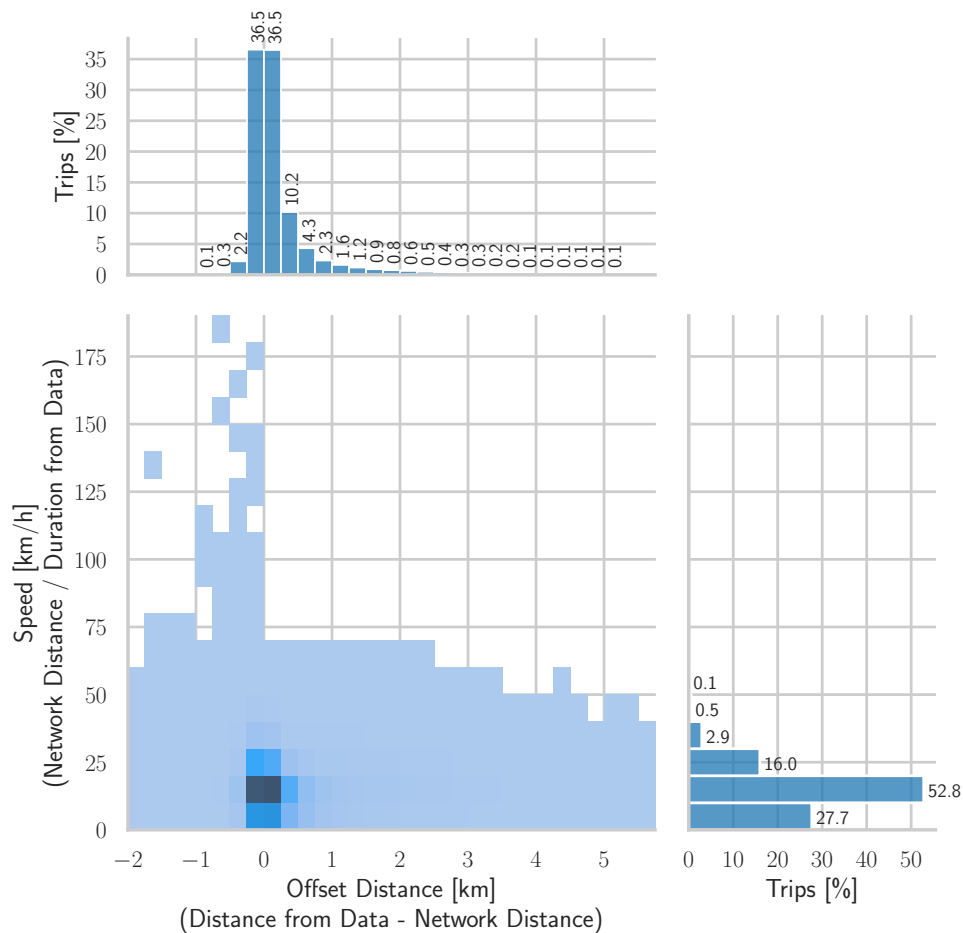
Figure 3.9.: The difference between the trip distances as available from the data and calculated from the OSM network used (the shortest path distances are calculated using trip origins and destinations). The speed on the y-axis shows the speed calculated using the shortest path distance from the network and the recorded travel duration from the data. The data refers to Manhattan trips from 6 to 7 June 2016.

than the distance calculated using the nearest network nodes. Interestingly, most of the trips have negative offset distances of less than 250 m, and all are below 1 km, indicating that using the closest network node did not induce a large error in calculations.

Besides the offset distances, it is also essential to consider if the travel duration is feasible according to OSM routes. As mentioned in section 3.2.2, the dissertation filters the possible false trips by only considering trips that have speeds between 1 mph (1.61 km/h) and 55 mph (88.5 km/h). If shortest-distance routes are valid for historical trips, then travel times should also be practical according to the distances of the shortest route. This is especially important for trips with negative offset distance; the reduction in travel distance could make some trips have impractically large speeds. Thus, Figure 3.9 also shows the speeds calculated using the distance from the network and travel time from the data. Some trips with negative offset distance have speeds much higher than the threshold of 88.5 km/h. However, such trips constitute an extremely small proportion of the overall data: only 64 out of 920465. These trips are also removed from the data set before solving the travel time scaling problem.
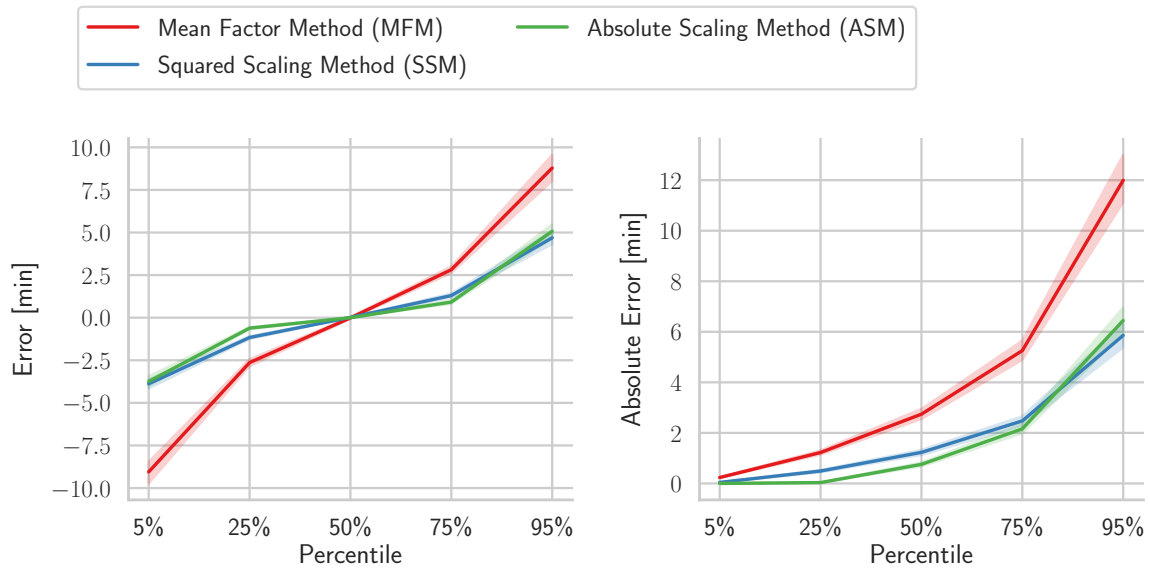
Figure 3.10.: The difference between the travel time from data and the scaled OSM network (error) for the small network (trips with origin and destination in Manhattan) using ΔT_{scale} of 30 minutes and NYC data from 6 to 7 June 2016.

The above discussion shows that the shortest routes of the network can provide the potential route used by historical trips. These paths contain the list of network edges that the dissertation uses in conjunction with the historical travel times to scale the network periodically. If the scaling method can scale the edges such that the travel times of the network are similar to the historical travel times, it ensures that the results of any AMoD algorithm tested on the network are also reliable.

3.3.4. Comparison of Scaling Errors

This section analyses the scaling errors for the Manhattan network. Figure 3.10 compares the scaling error of the three methods using ΔT_{scale} of 30 minutes and data from 6 to 7 June 2016. For the sake of conciseness, it combines the temporal scaling error from two days into one figure by first calculating the percentiles of the error for each ΔT_{scale} period, and then plotting them together. Thus, the shadow in the plotted lines represents the 95% confidence interval for each of the percentiles.

Figure 3.10 shows that both SSM and ASM provide more accurate travel time estimation than MFM method. Since MFM multiplies all the OSM edges by a mean factor, it produces symmetric positive and negative errors, leading to a mean error close to zero. However, in terms of individual trips, the MFM leads to higher scaling error: more than 50% of the trips have an absolute error of 2.74 minutes or more (50% percentile of 2.74 minutes), while 25% of the trips have an absolute error of more than 5.26 minutes (indicated from 75% percentile of 5.26 minutes), reaching to almost 12 minutes or more for 5% of the trips (indicated by 95% percentile of 12 minutes). In contrast, 50% of the trips in both SSM and ASM have an absolute error of less than 1.23 and 0.76 minutes, respectively. Even higher percentiles of the absolute error for SSM and ASM are almost half of the MFM: a 75% percentile of 2.47 and 2.16 minutes (in contrast to 5.26 minutes for MFM), and a 95% percentile of 5.86 and 6.45 minutes for SSM and ASM (in contrast to 11.99 minutes for MFM), respectively. Table 3.1 shows the percentile and maximum values of the absolute error.

Figure 3.11 shows the percentages of total trips within various ranges of the absolute error. First,

Scaling Method	Percentile					Maximum
	5%	25%	50%	75%	95%	
Mean Factor Method (MFM)	0.23	1.22	2.74	5.26	11.99	71.14
Squared Scaling Method (SSM)	0.04	0.49	1.23	2.47	5.86	59.66
Absolute Scaling Method (ASM)	0	0.03	0.76	2.16	6.45	69.66

Table 3.1.: Mean of absolute errors for data from 6 to 7 June 2016 and smaller Manhattan network and ΔT_{scale} of 30 minutes. The values are in minutes.
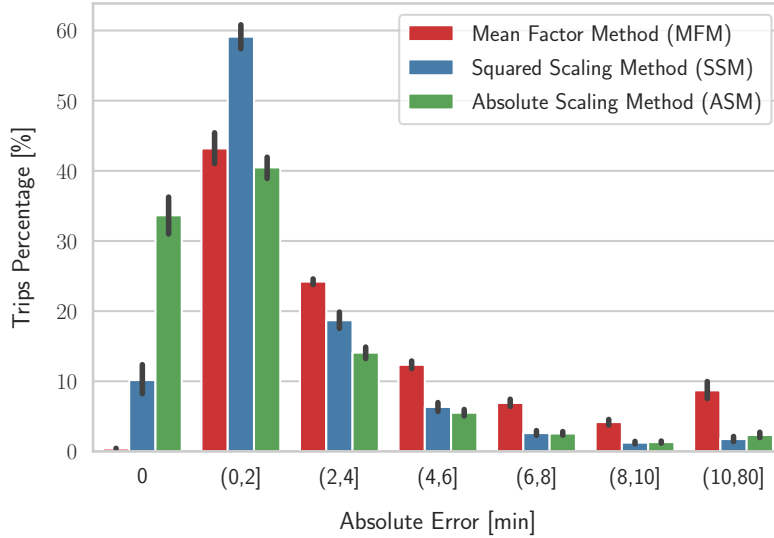

Figure 3.11.: Percentages of trips in different ranges of absolute error. An absolute error of less than 1 second is marked a 0.

MFM does not produce a significant proportion of trips with exact accuracy (an absolute error of less than 1 second). On the other hand, due to the edge-based scaling in SSM and ASM, they can reproduce exact travel times for a higher proportion of trips. Figure 3.11 also demonstrates a major difference between SSM and ASM; because the SSM uses the square of the scaling error as the objective function, it prioritizes the minimization of high magnitude errors. Thus, SSM has a lower 95% percentile and maximum absolute error than ASM, as shown in Table 3.1. Similarly, it has a lower proportion of trips with an absolute error of more than 10 minutes compared to ASM. Conversely, the ASM uses the absolute value of the scaling error as the objective function. Thus, it prioritizes the minimization of scaling error for a higher number of trips. Figure 3.11 also demonstrates this behavior; it has a much higher proportion of trips in lower ranges of absolute error. Almost 33.7% of the trips have no scaling errors with ASM compared to 10.2% for SSM and only 0.42% for MFM.

Lastly, it is also important to note that the comparisons in this section concerned only the edges E_u that were part of the shortest distance routes obtained from the network. Both SSM and ASM also have a second step where the scaling factors are applied to the rest of the edges. While the scaling of E_u is essential for direct journeys from origin to destination of historical trips, scaling other edges is important for trips not explicitly present in the historical data, such as journeys to pick up customers and reposition idle vehicles. Nonetheless, an accurate scaling and detailed analysis of these edges are difficult due to the unavailability of traffic data. Thus, the basic assumption of both
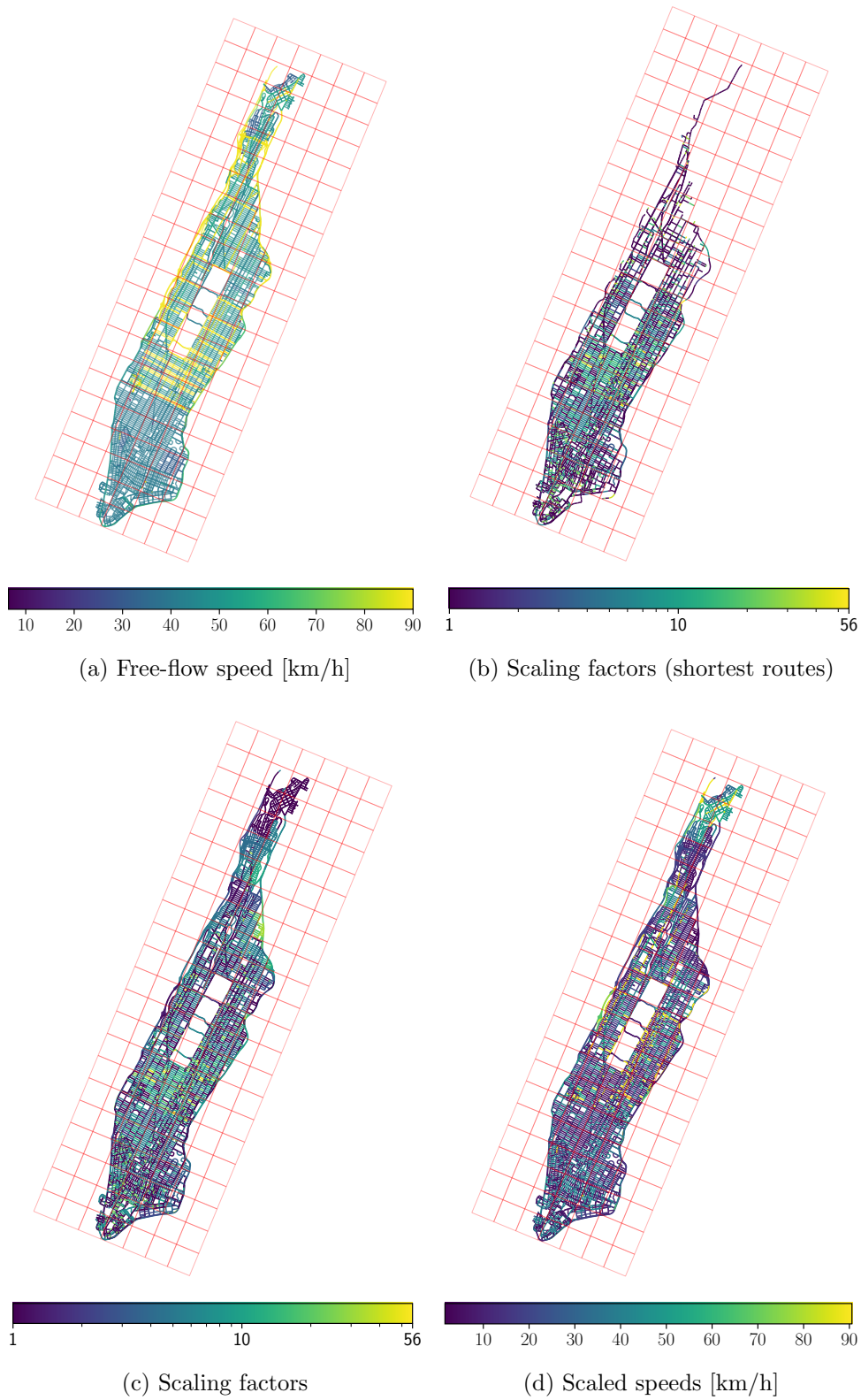
(a) Free-flow speed [km/h]

(b) Scaling factors (shortest routes)

(c) Scaling factors

(d) Scaled speeds [km/h]

Figure 3.12.: An example of scaling factors and speeds obtained using a regular grid with cell size Δs_{cell} of 1 km and ASM method. The data used is from 9 am to 9:30 am on 6 June 2016.

SSM and ASM is that the travel times should scale on a similar level in the surroundings of already scaled edges E_u, as described in section 3.3.2.

Fig. 3.12 illustrates an example of the ASM scaling for the rush hour (9 am to 9:30 am) of Monday morning. First, as shown in Fig. 3.12c, the scaling factor varies for different regions and links in the network. Second, as shown in Fig. 3.12b, the shortest-path routes do not necessarily include all of the network edges; rather, the areas with the highest number of origins or destinations (e.g., central Manhattan for the rush hour) provide the highest level of information for the scaling method. This is because the scaling is done in intervals of ΔT_{scale} (30 minutes) and only the trips within ΔT_{scale} are used for scaling, and thus, some regions may contain very few edges from E_u while others may not have any edge from E_u at all (refer to the cells in the north in Fig. 3.12b). Therefore, the scaling accuracy of a region also depends on the number of historical trips and shortest path edge samples available for a particular region. This is apparent by comparing the regions in Fig. 3.12b that have a small number of edge samples to the edges in Fig. 3.12c and the resulting scaled speed in Fig. 3.12d. However, this also depends on the time of the day as the pattern of trip origins and destinations change from morning to evening and during night hours. Nevertheless, both ASM and SSM methods compensate for the absence of the data by assigning the average scaling factor of a cell to the links without a travel time sample. This assumes that the traffic situation within a small cell should be similar for all the links in that specific cell.

3.3.5. Selection of Scaling Method

The above comparisons show that both SSM and ASM replicate historical travel times significantly better than MFM. SSM suits better to minimize the higher scaling errors due to the squared objective function; it has lower percentages of trips with scaling error in the higher range (10 minutes or more) than ASM. On the other hand, due to the absolute value function in the objective, ASM does not prioritize minimizing the scaling errors of higher magnitudes instead it tries to minimize the scaling error for a higher number of trips. Thus, the proportion of trips without any scaling error is much higher for ASM than SSM. Nevertheless, the difference in trip percentages of SSM and ASM for higher ranges of scaling errors are insignificant. On the other hand, ASM provides accurate travel times for a much higher percentage of trips than SSM. Therefore, the dissertation uses ASM for simulations with grid cell size Δs_{cell} of 1 km and scaling period ΔT_{scale} of 30 minutes.

3.4. Conclusion

The chapter presented the AMoD service types modeled in the dissertation along with the basic AMoD fleet management problem. The operation of AMoD fleet is controlled by the central FC located at AMoD operator. Two different VCO methods were presented in the chapter. The agent-based simulation framework that will be used in the rest of the dissertation was also presented. The chapter also presented the NYC taxi data and the city network used. To produce realistic travel times and reliable simulation results, the chapter also presented a travel time scaling method using historical trip data. Out of the discussed methods, ASM scaling method will be used for network scaling in the rest of the dissertation.

Chapter 4.

A Spatiotemporal Metric for AMoD Service Quality

The operation area of an AMoD service is often divided into regions for various reasons. On the one hand, these reasons could be purely analytical without directly impacting the AMoD operation, for example, to observe the patterns of origins and destinations in historical data or to spatially aggregate locations for privacy reasons before publication. On the other hand, many times, the AMoD operator would do such a division to improve or maintain a seamless service, for example, to estimate new depots that would suffice local demand or to regularly reposition AMoD fleet to meet the local demand. The dissertation concerns the latter category, i.e., the division into regions for maintenance purposes. In this regard, the primary assumption found in literature is that the individual regions are independent such that an AMoD vehicle can only serve the customers within that region. In contrast, this chapter presents spatial metric inspired by kernel density estimation (KDE) that links the whole AMoD operation area in the form of a heat map. The chapter shows that such a spatial metric can describe the AMoD service quality offered at individual points in the operation area. This spatial metric is used in the later chapters to improve the AMoD operation.

The chapter is organized as follows. Section 4.1 first describes the usual approach found in the literature for dividing the operation area into regions. Section 4.2 then describes the general KDE. It also describes the challenges faced when using it as a spatial metric for AMoD service quality and the changes required to adapt it for AMoD services. Finally, section 4.3 studies the effectiveness of the introduced spatial metric to describe the service quality offered to each customer. For this purpose, it uses the agent-based simulation and the NYC data introduced in the previous chapter.

4.1. Independent Operational Regions

Several researchers divide the AMoD operation area into a disjoint set of smaller regions [Dandl, M. Hyland, et al., 2019; Dandl, M. Hyland, et al., 2020; S. Hörl et al., 2019; Ruch et al., 2018]. From the perspective of AMoD service, the main application of such a division is the periodic repositioning of idle vehicles to areas of higher demand. This requires selecting a group of idle vehicles and relocating them to various locations. Ideally, such a rebalancing would perform best if applied on individual coordinates rather than the regional level. However, first, such a granularity level makes the solution space intractable. Second, the forecast of future supply and demand is not available on such a fine level. Therefore, the division into regions has been the preferred method of choice in literature [Dandl, M. Hyland, et al., 2020; Fagnant and Kockelman, 2014].

Multiple researchers have used the regional division with various repositioning methods. However, a major drawback in most of these works has been developing a consistent relationship among regions. Many works fundamentally assume that only the locations within a single region can interact with each other, i.e., a vehicle within a single region can only serve the customers with pickup points within

that region. This assumption reduces the complexity and easily solves any operational problem. The dissertation refers to this approach as *independent regions*.

The above assumption is generally valid and helpful when analyzing past data for studying various trends or developing strategies for geographically distant regions. Most of the time, the concerned AMoD operation area is within a city and the individual regions are in close proximity such that the decisions taken for one region also impact others, i.e., depending on the proximity of individual regions, customers within one region can be served by the vehicles in other regions. There is a strong requirement for a spatial relationship that links multiple regions into a single formulation based on the proximity of individual regions and, consequently, can describe the overall AMoD service quality coherently. The following section introduces such a formulation which will be used throughout the dissertation.

4.2. Reachability Density based Relations for AMoD services

The independent regions approach divides the AMoD operational area into regions assuming that the vehicles within a region would only pick up customers within that region. Some works like [AL-KANJ et al., 2020; FAGNANT and KOCKELMAN, 2014] have used methods based on hierarchical regions that provide an alternative approach to independent regions. However, a major drawback of these approaches is that the inter-regional relations are strictly defined by the geometries of the regions, i.e., the vehicle is assumed to be available for all or none of the geographical locations of the nearby regions. Some methods address this problem using a time-based partial availability of the vehicles in the nearby regions [FAGNANT and KOCKELMAN, 2014]. Nevertheless, a coherent formulation is lacking where inter-regional relations could be flexibly defined independent of the regional shapes.

Given the above requirement, this section introduces a reachability density (RD) based distribution model for AMoD fleet. The formulation is inspired by kernel density estimation (KDE) which is a widely used technique in statistics. Fundamentally, instead of using the definition of regions to group the vehicles and customers spatially, the RD based formulation uses the reachable distances from a point to describe the fleet distribution. This leads to the following features of the introduced RD based formulation:

- **Independent of regional shapes**: The defined RD based spatial relations do not require defining a disjoint set of regions; rather, it naturally links the points in the operation area into a consistent spatial formulation.

- **Geographical heat map**: The formulation provides the probability that a new customer can be served by the AMoD fleet within the given time constraints. This allows the MSPs to map the offered service quality in the operational area as a geographical heat map.

- **Active usage in AMoD operation**: The formulation helps to measure the service quality offered at the current time, and thus, it can be used to actively improve the AMoD operation by incorporating it into operational decisions.

Due to the above characteristics, the RD based relations introduced in this section serve as an important spatial metric for measuring the AMoD service quality. The following sections describe the RD based relations in detail. The sections are structured as follows: section 4.2.1 first introduces the general KDE method used in statistics, section 4.2.2 describes how KDE can be adopted for AMoD service quality measurements, section 4.2.3 describes the RD formulations, and finally, section 4.2.4 introduces the methods used in the dissertation to calculate the maximum reachable distances from each node of the city network.

4.2.1. Kernel Density Estimation

The KDE is a well-known non-parametric probability density function (pdf) estimator first introduced by PARZEN [1962]. KDE tries to automatically adapt itself to the shape of the underlying density function. Let $x_1, x_2, ..., x_N \in \mathbb{R}^d$ be a set of independent data points drawn from a probability density function $p(x)$, then a KDE is calculated as

$$\hat{p}(x) = \frac{1}{N V_d^{(k)} h^d} \sum_{i=1}^{N} k(x, x_i, h) \tag{4.1}$$

$$k(x, x_i, h) = K\left(\frac{x - x_i}{h}\right) \tag{4.2}$$

where $K : \mathbb{R}^d \mapsto \mathbb{R}$ is a smooth kernel function, $h > 0$ is the bandwidth for smoothness and x_i is a data point. V_d^k is a kernel- and dimension-dependent normalization factor so that the integral $\int \hat{p}(x) dx = 1$ [X. WANG, 2005].

Many smooth kernels are generally used in KDE, e.g., triangular, Gaussian, triweights etc. However, as described in the following section, the most relevant kernel for the AMoD services is the simple triangular function, given as:

$$K(x) = \begin{cases} 1 - \|x\|^s, & \text{if } \|x\|^s \leq 1 \\ 0, & \text{otherwise} \end{cases} \tag{4.3}$$

The dissertation uses the Euclidean norm in the above equation, i.e. $s = 2$, for which $V_d^{(k)}$ is $\frac{\pi}{3}$ [X. WANG, 2005].

The fundamental usage of a KDE is to estimate the underlying probability density function (pdf) of a data set. The advantage of KDE is that it does not require making any assumption on the specific parametric family that the pdf might belong to; instead, it adapts the shape of estimated pdf directly from the data, making it a very useful tool for data drawn from complicated distributions. The most important parameter of a KDE, i.e., the bandwidth h, is usually calculated using various methods to minimize the asymptotic mean integrated square error (AMISE). The data centers and the overall spread of the estimated pdf are heavily dependent on the choice of bandwidth h (the spread of a single data point) [GRODZEVICH and ROMANKO, 2006].

4.2.2. Adopting Kernel Density Estimation for AMoD services

A KDE seamlessly combines the impacts of multiple data points through kernels. Therefore, a KDE-inspired spatial relation can potentially integrate the overall impact of the AMoD operation area into a single function. However, as described in the following, KDE requires certain modifications to achieve this objective. The modifications described in this section deal directly with individual coordinates, and thus, do not require defining operational regions. However, certain FC operations like repositioning would require using it together with operational regions which would require further adaption of the approach as described in the next chapter.

The most significant modification is in interpreting what a kernel bandwidth h represents. While in KDE it simply describes the spread of the kernel function, for the purpose of AMoD services, it is understood as an approximation of the maximum *reachable* (or *servable*) distance of a single vehicle (or customer request) within the maximum allowed customer waiting time ΔT_{max}. Thus, the kernel function is called the reachability function (RF) of a vehicle or customer request. An RF provides the probability of a vehicle serving a customer in its surrounding. On the one hand, this naturally

eliminates the fundamental limitation of the independent region-based approaches that treat each region separately without due regard to the proximity of other regions. On the other hand, unlike hierarchical relationship [Al-Kanj et al., 2020; Fagnant and Kockelman, 2014], the partial impacts and reachable points in the neighboring regions are not directly dependent on shapes of regions; instead, they depend on the bandwidth h.

The above interpretation causes some major inconsistencies if the traditional definition of KDE is used for AMoD services as described below:

1. The primary focus of the bandwidth selection algorithms in KDE is to find an h that would improve the estimate of the underlying pdf. In the case of geographical locations of customers and vehicles, it would only calculate an h that provides a good estimate of the underlying probability densities of the customer and vehicle locations. Such a value of h would not correspond with the above interpretation of kernel functions as RFs, i.e. the h would not correspond with the maximum reachable distance from a vehicle or customer location.

2. The customer and vehicle data in an AMoD scenario are dynamic, and thus, the underlying density of their location is expected to change throughout the day and week. Therefore, the value of h derived from bandwidth estimation algorithms can vary significantly depending on the time of the day. Any operational decision made on the basis of such a density function may lead to inconsistent long-term results.

3. Even for a static problem instance that aims to make an operational decision for all vehicles and customers simultaneously, the bandwidth h obtained using traditional KDE approaches might be significantly different for the customer and vehicle distributions. For example, consider a repositioning problem instance that balances the vehicle supply with customer demand. Even if the supply-demand balancing problem tries to close the gap between the separately calculated customer and vehicle KDEs, the vehicles still may not serve the intended customer because of different bandwidths. For example, consider the case when customer and vehicle KDE has a bandwidth of 500 m and 2 km, respectively. The algorithm will falsely balance very few vehicles assuming that the vehicles have a reachability of 2 km, which may not be applicable due to the average speed in the area of operation.

4. Since the KDE is normalized with the number of available data points (Eq. 4.1) to make the integral unity, the scales of customers and vehicles KDE might be significantly different. For example, consider the case with 100 expected customers and 300 idle vehicles with KDEs calculated using Eq. 4.1. For the use-case of repositioning, the algorithm would try to unnecessarily send a high number of vehicles near the expected customers as the customer KDE would be on a similar scale as idle vehicles KDE due to normalization.

To resolve issues 1 and 2, for each location x_i, the dissertation selects the bandwidth h_i according to the current network state (reachable distances considering current network state) and maximum waiting time ΔT_{max}. For issue 3, it is noticed that a major problem is caused by normalization of the KDE on the same scale such that $\int \hat{p} dx = 1$ irrespective of the number points, which is usually required because KDE is estimating a pdf. However, for this dissertation, estimating a pdf is not as important as using a consistent scale for customers and vehicles. Thus Eq. 4.1 is modified as:

$$\bar{p}(x) = \frac{3}{\pi} \sum_{i=1}^{N} \frac{k(x, x_i, h_i)}{h_i^2} \tag{4.4}$$

where k is the triangular kernel or RF given in Eq.4.3. Even though any kernel function (guassian, triweights etc) can be used as a RF, the dissertation uses triangular function due to its simpler interpretation: for a vehicle located at the center of an RF, the probability of serving a customer is highest when the customer is located at the center of the RF and it linearly decreases as the euclidean distance from the center increases. Similarly, from the perspective of a customer located at the center of an RF, the probability of being served by a vehicle is the highest when the vehicle is located at the center and it linearly decreases as the euclidean distance from the center increases.

Due to the above modification, Eq. 4.4 does not represent a pdf, however, it has following useful properties:

- The usage of RFs causes the impact of single data points to spread to their surroundings based on the incorporation of information from three important sources: 1) underlying network 2) geography of the operation area and 3) the AMoD service model. First, the underlying network is incorporated by calculating the maximum reachable distance h_i for a data point located at x_i using the city network and its current state. Second, the geography of the operation area plays a part in determining the impact of a data point as the distance from its center increases. In this regard, the dissertation uses 2D triangular function. However, the formulation can also use a more complex geometrical function. Third, the AMoD service model is incorporated by using δT_{max} as a criterion to calculate the maximum reachable distance h_i. This is especially relevant for the AMoD service model studied in the dissertation, i.e. AMoD without pooling of customer requests; the pooling of customers makes it more complicated to calculate the h_i as the reachability and the ability to serve new customers additionally depends on the remaining capacity in the vehicle and the detour time of the onboard customers.

- The integral of Eq. 4.4 over the AMoD operation area is equal to the number of data points, i.e. $\int \bar{p}(x)dx = N$. This indicates that the formulation not only spreads the influences of a single data point using the RF but also coherently combines the spatial influences of all RFs such that the integral represents the total number of data points.

- The above property further leads to two useful properties of the formulation. The first one is in terms of it being a spatial metric for *reachability* from data points: a higher value of $\bar{p}(x)$ for any geographical location x means that x is under the influence of a higher number data points, and therefore, there is a higher reachability to (or from) the nearby data points. It should also be noted that while multiple data points influence x, the influence of the closest data point will be the highest due to triangular RF. Thus, Eq. 4.4 coherently combines the influence of single and multiple data points. The second useful property is in terms of consistency across problems with different data sizes; even if two separate problems have different numbers of total data points, the values of $\bar{p}(x)$ for a location x can still be compared to measure the reachability. For example, if problem 1 has 100 data points located at certain locations and problem 2 has 500 data points located at exactly the same locations as problem 1 (overlapping data points), then the value of $\bar{p}(x)$ for a location x will be simply higher (or at least equal) for problem 2 than problem 1 due to the access to a higher number of nearby data points. This property is especially useful for AMoD services since the number of available vehicles and customers can be compared consistently across multiple VCO batches.

Because of the characteristics, the dissertation terms Eq. 4.4 as reachability density (RD)— *reachability* because the values of $\bar{p}(x)$ do not represent probability density like in KDE, rather it represents the likelihood of reaching to (from) the nearby data points from (to) a location x. If it is calculated for available vehicles, then it represents the reachability density of supply, i.e. the ability of vehicles
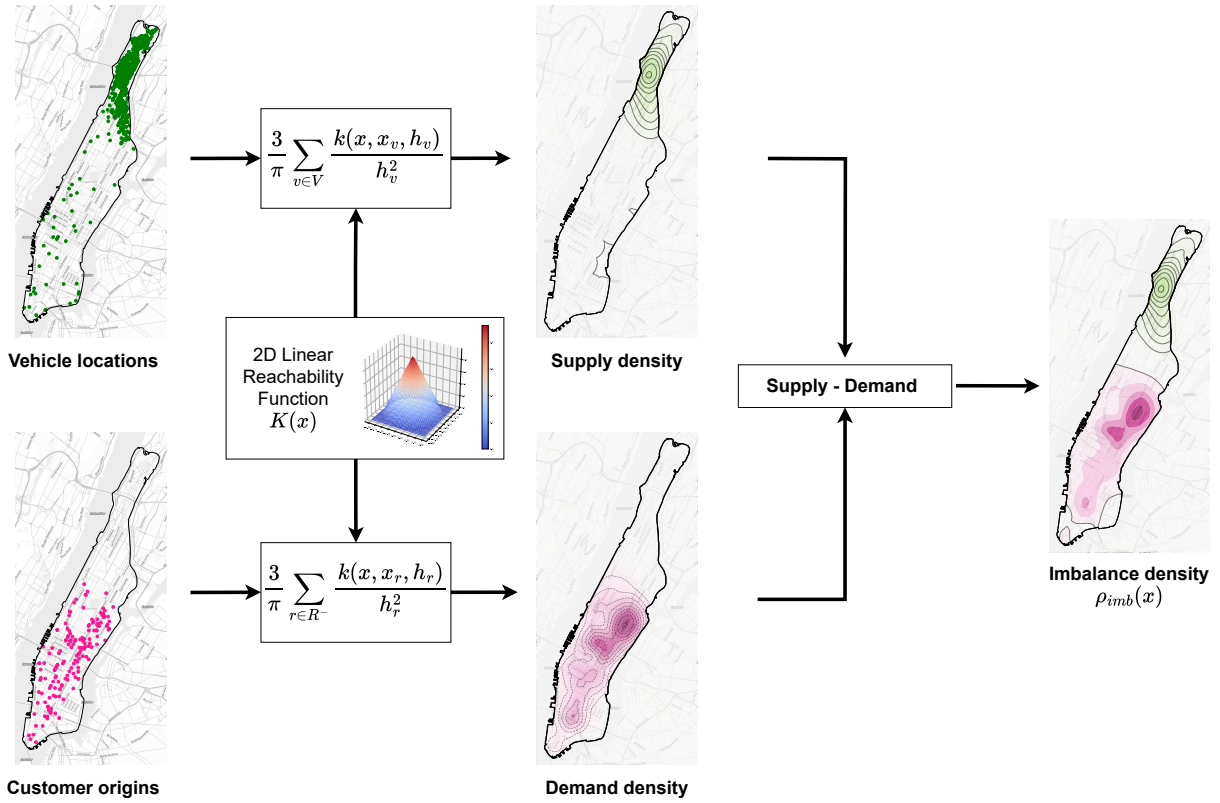
Figure 4.1.: Flow chart for calculating imbalance density (ID)

to reach a certain point within ΔT_{max}; the higher value of $\bar{p}(x)$ for a geographical location x, the higher the chances are that the nearby vehicles can reach this point within ΔT_{max}. Similarly, if it is calculated for customer pickup locations, then it represents the reachability density of demand, i.e. the potential to reach nearby customers from a point x; the higher value $\bar{p}(x)$ is for a location x, the higher the chances are that a vehicle located at x will be able to serve a customer.

4.2.3. Reachability Density based Metric for Supply-Demand imbalance

The RD based spatial relation presented in the previous section built a relationship between a dispersed set of locations, either the locations of vehicles (supply) or customer pickup locations (demand). However, from the perspective of AMoD operations, a more useful parameter is to measure the supply-demand imbalance in the operational area than individual RD of supply or demand. As described in the previous section, one of the characteristics of RD is the consistency for problems of different data sizes. Because of this property, the supply and demand densities could be subtracted from each other to get a RD function for supply-demand imbalance, termed as imbalance density (ID) in the dissertation. The following describes the calculation procedure for ID.

Let V^+ be the set of current locations of available vehicles and R^- be the set of customer pickup locations. Then, using Eq. 4.4 the 2D density function for ID can be obtained by subtracting the supply density from the demand density, given as:

$$\rho_{imb}(x) = \frac{3}{\pi} \left(\underbrace{\sum_{v \in V} \frac{k(x, x_v, h_v)}{h_v^2}}_{\text{supply density}} - \underbrace{\sum_{r \in R^-} \frac{k(x, x_r, h_r)}{h_r^2}}_{\text{demand density}} \right) \tag{4.5}$$

where x_v represents the location of the vehicle v and h_v the maximum reachable distance from x_v. Similarly, x_r represents the requested pickup location of the customer r and h_r the maximum reachable distance from x_r. The values of h_v and h_r depend on the city network and its current state as described in the previous section.

The dissertation proposes that the function $\rho_{imb}(x)$ indicates two important observations. First, since it is the difference of supply and demand density, a positive value of $\rho_{imb}(x)$ would imply a surplus of vehicles and a higher probability that the customers around the location x would be served by the AMoD service. Second, because Eq. 4.5 uses triangular kernels with bandwidths dependent on reachable distances, a higher positive value of $\rho_{imb}(x)$ would mean a lower limit on possible pickup delay. However, both of the above observations depend on a couple of factors. First, it depends on how accurately h_v and h_r represent the actual reachable distances within ΔT_{max}. Second, the dissertation uses euclidean norm with the triangular function as RF (Eq. 4.3), which means that the RF assumes a strict circular boundary for the maximum reachable distances. However, the real networks of a city are complex and the real maximum reachable distances may not have such a symmetrical limit on reachable distances in all directions.

Third, the actual assignment of vehicles to customers is done inside the VCO, which may not necessarily correspond to the primary assumptions of the triangular RF uses, i.e., an AMoD vehicle prefers the customers located near its location and the preference decreases as the euclidean distance from the location of the vehicle increases. Nonetheless, this assumes that VCO tries to assign vehicles to customers such that the pickup distances are also minimized. This could be due to a direct minimization of the pickup distance [M. HYLAND and MAHMASSANI, 2018] or as a result of an indirect minimization of a secondary goal with a similar outcome such as the minimization of pickup travel time [ALONSO-MORA, SAMARANAYAKE, et al., 2017; ALONSO-MORA, WALLAR, et al., 2017] or traveling cost for the pickup journey [DANDL, M. HYLAND, et al., 2020].

4.2.4. Calculation of Reachability Function Bandwidth

In contrast to [A. A. SYED, DANDL, KALTENHÄUSER, et al., 2021], where a fixed reachability bandwidth is used for all kernels, the dissertation calculates the bandwidths according to the current state of the dynamic network. Let the city network be represented by the directed graph $G_{net}(N_{net}, E_{net})$. The dissertation uses only the locations of nodes N_{net} for the calculation of bandwidths; the intermediary points on the network edges are not taken into account for maximum reachable distance. In reality, an estimation that includes network edges can provide a better estimation of the reachable distances within ΔT_{max}, however, for the purposes of the dissertation, restricting to network nodes suffices since the customer locations are also explicitly mapped to the nearest city nodes (section 3.2.1). So a vehicle journey could only start or finish at one of these network nodes. The following describes the method used for calculating the reachability bandwidths.

The method depends on two main components: 1) the nodes reachable within a travel time of ΔT_{max} and 2) the euclidean distances of nodes. For each source node $n_s \in N_{net}$ (representing the center of the RF), all the nodes that are reachable within the traveling time ΔT_{max} are first filtered, represented by the set $N_s \subset N_{net}$. Second, the euclidean distances from n_s to the nodes in N_s are calculated, which are used to derive the RF bandwidths.

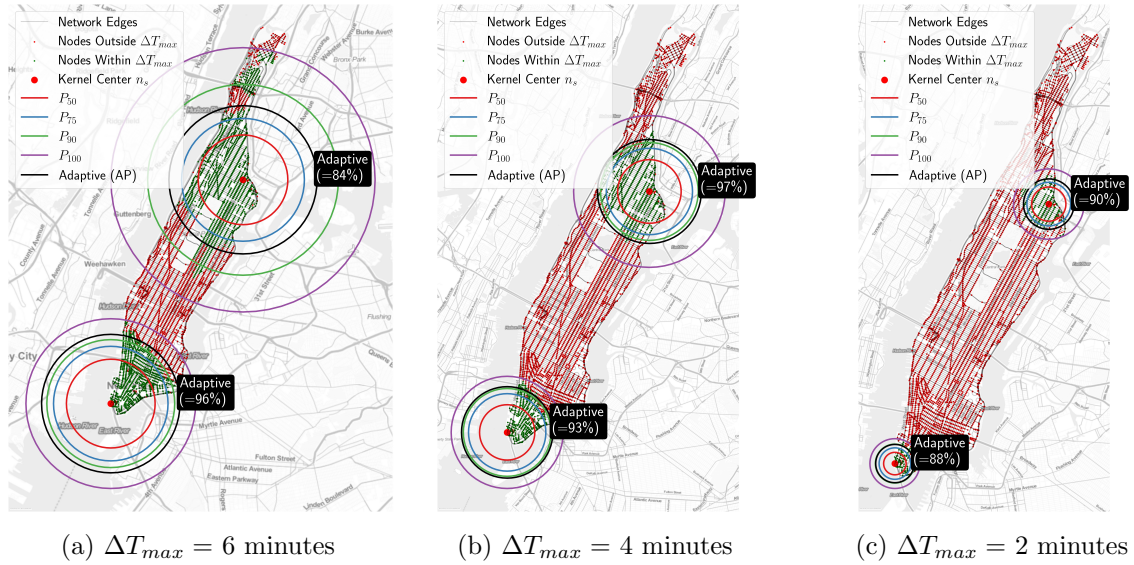(a) $\Delta T_{max} = 6$ minutes (b) $\Delta T_{max} = 4$ minutes (c) $\Delta T_{max} = 2$ minutes

Figure 4.2.: An example of RF bandwidths using free-flow network speeds. The radius of P_i corresponds to the radius that makes ith percentile of nodes around the origin node to be reachable within travel time of ΔT_{max}

When selecting the RF bandwidths, a problem is faced; the city network is usually non-homogeneous in different directions, and thus, the euclidean distances of the farthest reachable node might significantly vary in different directions, as shown in Figure 4.2a. Moreover, in some cases, there could even be a group of non-reachable nodes intervening with a group of reachable nodes; Figure 4.2a shows such an intervening group in the north of Manhattan due to the unavailability of direct access to the intervening group and a highway (Harlem River Dr) in the north-east part making the farther north part of Manhattan reachable within ΔT_{max}.

In contrast, the triangular RF assumes a symmetric reachability boundary around the source node n_s. Thus, the question arises as to which directional distance should be chosen for approximating the RF bandwidth or (equivalently for the dissertation) which node within N_s should be selected for approximating the reachable distance around n_s. If the euclidean distance of the farthest node in N_s is selected, the RF will wrongfully assume that all the nodes within that radius are reachable within ΔT_{max} — a significant concern for the nodes in other directions. Figure 4.2a also highlights this situation whereby using euclidean distance of the farthest node, many unreachable nodes in the center of Manhattan wrongfully appear reachable to the RF. To resolve this issue, the dissertation suggests two methods: 1) the fixed percentile (FP) method and 2) the adaptive percentile (AP) method.

First, instead of directly dealing with the distances of N_s to determine the bandwidth, the FP uses percentiles of the distances of N_s to guarantee that a certain percentage of N_s are within the selected radius, represented by P_i for ith percentile. For example, for a node n_s, a bandwidth of $h = P_{75}$ would guarantee that the 75% of the nodes in N_s are within the RF. Figure 4.2 shows examples of multiple P_i for different values of ΔT_{max}. When calculating P_i, if the desired P_i falls between two euclidean distances, P_i takes the lower value instead of interpolating the distance. FP uses the same percentile i while calculating the bandwidths of all the nodes as indicated by the method's name. Algorithm 1 summarizes the FP method.

In the FP method, the right choice of P_i plays a vital role in excluding the unreachable nodes inside

---

Algorithm 1 The FP method for calculating RF bandwidths.

Input: The city network graph $G_{net}(N_{net}, E_{net})$
       The required percentile i

Output: A dictionary M of maximum distances

 1: $M \leftarrow$ empty dictionary
 2: **for** source node n_s in N_{net} **do**
 3:     $N_s \leftarrow$ Calculate the nodes in $N_{net} \setminus \{n_s\}$ that have travel time less than ΔT_{max} from n_s
 4:     $D \leftarrow$ Calculate array of Euclidean distances from n_s to nodes in N_s
 5:     $d \leftarrow$ Calculate the ith percentile of D and take the lower value if ith percentile fall between two values of D
 6:     $M(n_s) \leftarrow d$
 7: **end for**

Algorithm 2 The AP algorithm for calculating RF bandwidths.

Input: The city network graph $G_{net}(N_{net}, E_{net})$
       The minimum percentage of reachable nodes q
       The minimum percentiles i_{min}

Output: A dictionary M of maximum distances

 1: $M \leftarrow$ empty dictionary
 2: **for** source node n_s in N_{net} **do**
 3:     $N_s \leftarrow$ Calculate the set of nodes in $N_{net} \setminus \{n_s\}$ that have travel time less than or equal to ΔT_{max} from n_s
 4:     $D \leftarrow$ Calculate array of Euclidean distances from n_s to nodes in N_s
 5:     $i \leftarrow 100$
 6:     **while** $i > i_{min}$ **do**
 7:         $d \leftarrow$ Calculate the ith percentile of D and take the lower value if ith percentile fall between two values of D
 8:         $N_d \leftarrow$ Calculate the set of nodes within the radius d of n_s
 9:         $N_{reach} \leftarrow$ Calculate the set of nodes in $N_d \setminus \{n_s\}$ that have travel time less than or equal to ΔT_{max} from n_s
10:         $ratio \leftarrow 100 \times N_{reach}/N_d$
11:         $M[n_s] \leftarrow d$ $i \leftarrow i - 1$
12:         **if** $ratio \geq q$ **then**
13:            break
14:         **end if**
15:     **end while**
16: **end for**

---

the RF; for example, for the node marked in the south of Manhattan in Figure 4.2a, the majority of unreachable nodes are excluded from the RF using P_{90}. However, the usage of the same P_i for all the nodes may cause problems, especially for the anomalous nodes; for example, for the node marked in the north of Manhattan in Figure 4.2a, a large number of unreachable nodes are included in the RF using P_{90} in contrast to the node in the south.

While the FP method selects the bandwidth based on how many N_s nodes are covered by the RF, the P_i used still requires manual adjustments. Secondly, as mentioned above, the same P_i may not work for different parts of the city due to a discontinuous group of reachable nodes.

Therefore, the AP method further modifies the FP method to include the number of unreachable nodes and individually adjust the used P_i for each n_s. It starts with P_{100} and iteratively reduces the used percentile i till the percentage of the reachable nodes within P_i is greater than or equal to q or a minimum percentile limit I_{min} is reached. The dissertation uses the convention that a particular instance of AP with a specific value for q and i_{min} is represented as $AP_{(q,i_{min})}$. As shown in Figure 4.2, the AP uses different percentiles for each node of n_s and automatically adjusts the used percentiles to avoid including too many unreachable nodes irrespective of the used ΔT_{max}. The method is described in Algorithm 2. The dissertation uses AP method to calculate RF bandwidths.

4.3. Case Study for Spatiotemporal Relations

This section uses the simulation setup described in chapter 3 to evaluate the effectiveness of ID based spatial relations. Since the main objective of the dissertation is to use these relations in order to improve the AMoD operation, the section uses the agent-based simulation to study two fundamental hypotheses: (1) the AMoD vehicles are not strictly bound within a certain region for serving the customers when the regions are small and contiguous and (2) an RD based relation for vehicle supply and demand (i.e. ID) better explains the service quality offered to individual customers than the assumption of independent regions. The section uses the customer waiting time and the probability of being served as the main criteria for measuring service quality.

Since the dissertation uses variable travel times (section 3.3.5) which can significantly affect the reachable distances — and the used RF bandwidths for enroute vehicles after each ΔT_{scale} —, the section additionally uses constant travel times based on free-flow (FF) speed for detailed analysis of the above assumptions, i.e. FF and $2\times$FF travel times.

The ID is calculated for each of the VCO batches as follows. The simulation stores the locations of all available vehicles and customers considered in each VCO batch. These are later utilized in a post-processing stage where for each VCO batch, the locations of the considered vehicles and the customers are used to calculate the ID using Eq. 4.5. The goal here is to analyze if the ID provides a good spatial metric to measure the offered AMoD service quality; the section does not focus on improving the AMoD operations using the ID function which is studied in the following chapters.

The section uses an AMoD fleet of 3000 vehicles and ΔT_{max} of 6 minutes. For the VCO the parameters in Eq. 3.3 are kept fixed, given as (ζ, f^D, c^D, c^F)=(2.5\$/served customer, 0.5\$/km, 0.25\$/km, 25\$/vehicle per day). Table 4.1 lists the default values of the parameters used in the current chapter.

4.3.1. Comparison Vehicle Control Optimization Methods

The section first compares the performances of the two VCO methods. As shown by the percentage of customers served ($S_\%$) in Figure 4.3, the AMoD service can serve almost all of the customers with a fleet of 3000 vehicles and free-flow speed even when only idle vehicles are considered for

Parameter / Method	Symbol	Default Value / Strategy
Base fare	ζ	$2.5 per customer
Distance based variable fare	f^D	$0.5 per km
Distance based cost	c^D	$0.25 per km
Fixed maintenance cost of vehicle	c^F	$25 per vehicle per day
Maximum allowed waiting time of customers	ΔT_{max}	6 minutes
The time period used for city network scaling	ΔT_{scale}	30 minutes
Travel time scaling method used	—	ASM (section 3.3)
RF bandwidth calculation method	—	$AP_{90,90}$ (section 4.2.4)
Fleet size	—	3000 AVs
Vehicle assignment method	VCO	The current chapter compares VCO_{idle} and $\text{VCO}_{enroute}$
Batching period	ΔT_{batch}	30 seconds

Table 4.1.: The default simulation configuration used in chapter 4.

assignment (VCO_{idle}). The main reason for this is that with free-flow travel times, the vehicles can reach larger areas within ΔT_{max} and serve the assigned customers faster. As the city network gets slower, the $S_\%$ of VCO_{idle} drastically decreases from a value of 98.1% to just 22.9% for the realistic travel times using ASM method, causing a decrease in overall profit. Additionally, with increased travel times, the mean waiting time W_{mean} and mean pickup distance D_{mean}^{pick} are also increased for VCO_{idle} method.

In regards to the consideration of the enroute vehicles for assignment, two general traits of the $\text{VCO}_{enroute}$ could be observed: lower D_{mean}^{pick} and higher W_{mean}. Including enroute vehicles increases the number of vehicles that could be assigned to customers. Since the VCO only considers the distance-dependent cost in the objective function (c^D in Eq. 3.3a), any vehicle that is closer to the pickup location of the customers (idle or enroute) is assigned to the customer regardless of the waiting time. For the free-flow speed, this causes a degraded performance of $\text{VCO}_{enroute}$ over VCO_{idle}; there is enough supply of idle vehicles within the travel time range of ΔT_{max} due to large vehicle outreach, but the $\text{VCO}_{enroute}$ preferred assigning enroute vehicles to customers due to shorter pickup distance from the last point in their paths. As the network travel times become more realistic, the inclusion of enroute vehicles in the $\text{VCO}_{enroute}$ produces significantly higher profit than VCO_{idle}: the $S_\%$ increased from 22.5% to 35.1% and the profit increased from 0.31 million to 0.52 million USD for ASM, however, W_{mean} also increased from 3.0 minutes to 4.8 minutes.

Figure 4.4 shows the performance on the temporal axis. The first thing to notice is the performance degradation with the passage of simulation time; without repositioning of idle vehicles, the AMoD fleet starts to accumulate in different parts of the city, causing regional supply-demand imbalances. Depending on the city network characteristics and the pattern of customer origins and destinations, this supply-demand imbalance can worsen over time, leading to degraded performance. Since the free-flow speed allows the AMoD fleet to have a significantly large outreach, the effects of supply-demand imbalances are least noticeable for free-flow speed. On the contrary, since the customers can be served quickly with the free-flow speed, the fleet utilization ($U_\%$) remains relatively low during the morning and evening peak hours even though $S_\%$ remains high. This indicates that with free-flow speed the AMoD fleet had the capacity to serve an even higher number of customers.

The impact of fleet accumulation is most noticeable for the ASM method. At the beginning of
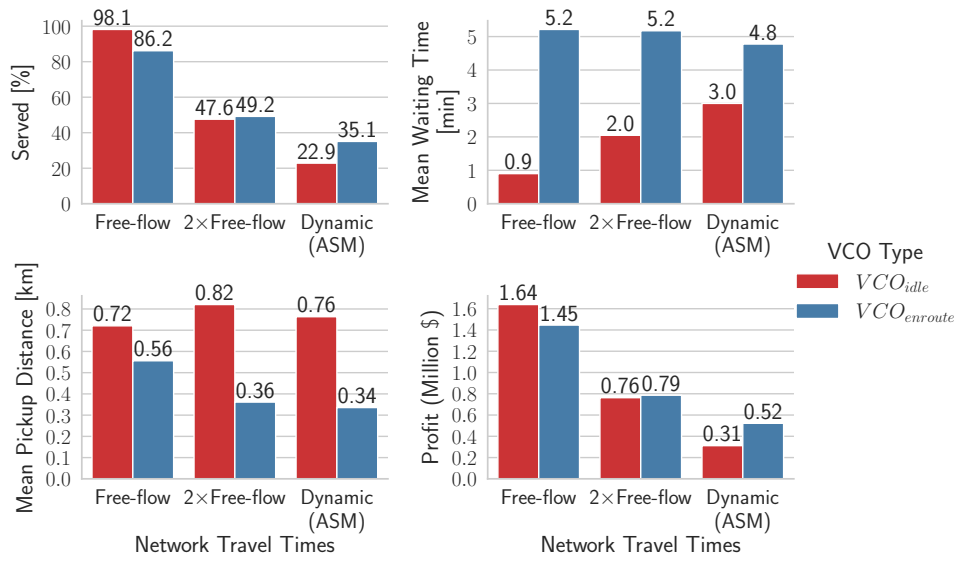
Figure 4.3.: The comparison of VCO$_{idle}$ and VCO$_{enroute}$ using a fleet of 3000 vehicles.

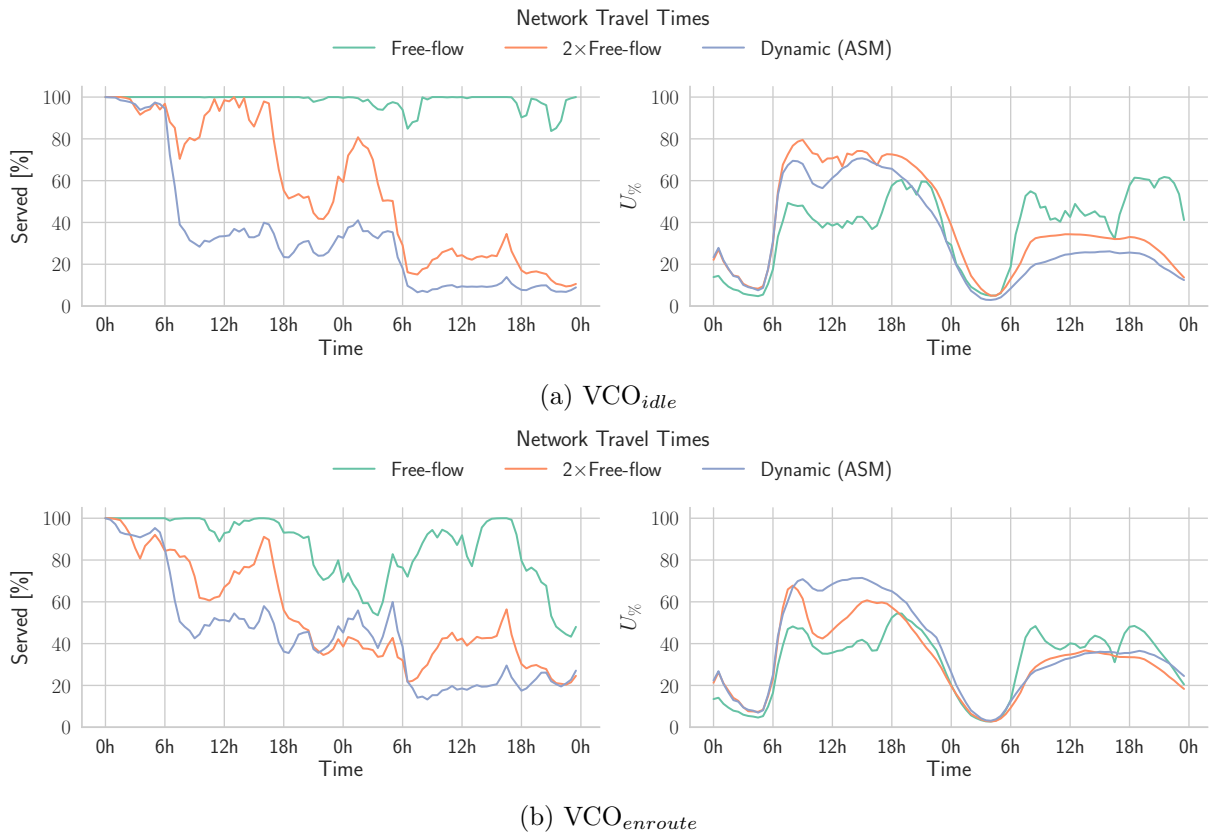(a) VCO$_{idle}$

(b) VCO$_{enroute}$

Figure 4.4.: The temporal performance with 3000 vehicles and ΔT_{max} of 6 minutes. The values are calculated using an aggregation period of 30 minutes.
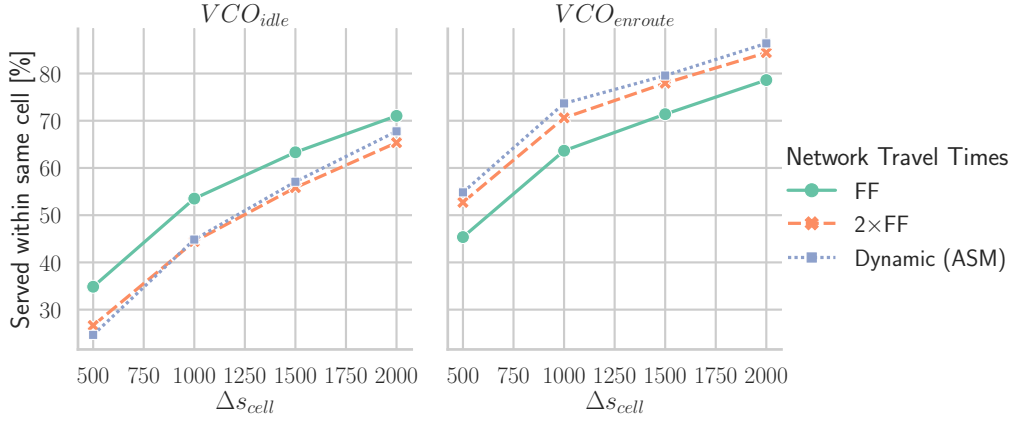
Figure 4.5.: Percentages of total served customers assigned to vehicles from the same zone as the customer pickup location

the simulation, with a better distribution of AMoD fleet and lower customer demand (0 to 6h of the first simulation day), both VCO methods can serve most of the customers with ASM travel times. However, during the day time when the network is slower and the customer demand is high, AMoD service is only able to serve less than 40% for VCO$_{idle}$ and 60% for VCO$_{enroute}$ of customer demand. On the second day of the simulation, most of the AMoD fleet remains accumulated in different parts of the city, forcing the AMoD service to have even lower $S_\%$ and $U_\%$ on the second day of simulation. DANDL, M. HYLAND, et al. [2020] also observed the same phenomenon, where, in a whole week of simulation, the $U_\%$ consistently dropped with each simulation day. The effects of accumulation are slightly lower for 2×free-flow speed than ASM method. This leads to the conclusion that the accumulation of AMoD fleet and its impact on performance increases with the stress on the city network. The accumulation of vehicles and the requirement of repositioning will be discussed in more detail in section 5.1.

Next, the section analyses how strictly the vehicles were bound to regions while serving the customers. It uses the regular grids with cell size Δs_{cell} defined in section 3.2.3. Since the VCO does not bound the assigned vehicles to be in the same region as the customer pickup location, the purpose here is to observe how often the vehicles crossover and serve the customers in other regions. Figure 4.5 shows the percentage of customers out of the total number of customers served that were assigned to vehicles in the same region as the customer pickup locations ($S_\%^{region}$). In general, for smaller regions, the vehicles are very loosely bound to serving customers from the same region; for Δs_{cell} of 500m and ASM travel times, $S_\%^{region}$ is only 24.7% for VCO$_{idle}$ and raises to 54.8% for VCO$_{enroute}$. With an increase in the size of regions Δs_{cell}, more customers are served from the same region. However, even for a Δs_{cell} of 2km, $S_\%^{region}$ is only 67.8% for VCO$_{idle}$ and 86.4% for VCO$_{enroute}$. For VCO$_{enroute}$, it should be noted that the availability location of vehicles is used for $S_\%^{region}$ and not their current locations. In reality, the AMoD operator might use clustering algorithms on past data to define regions that may better restrict the vehicles to specific regions. Nevertheless, the region sizes cannot be grown indefinitely, as it would raise additional challenges for FC methods since it will have to additionally decide where to position the vehicle within each region.

Figure 4.6 shows $S_\%^{region}$ for dynamic travel times and different values of ΔT_{max}. A lower ΔT_{max} prunes the solution space strictly such that the far-off vehicles have a lower probability of picking up the customers within the given ΔT_{max}. This leads to more vehicles being assigned to customers within the same region. The differences in $S_\%^{region}$ for varying ΔT_{max} is even higher for VCO$_{idle}$:
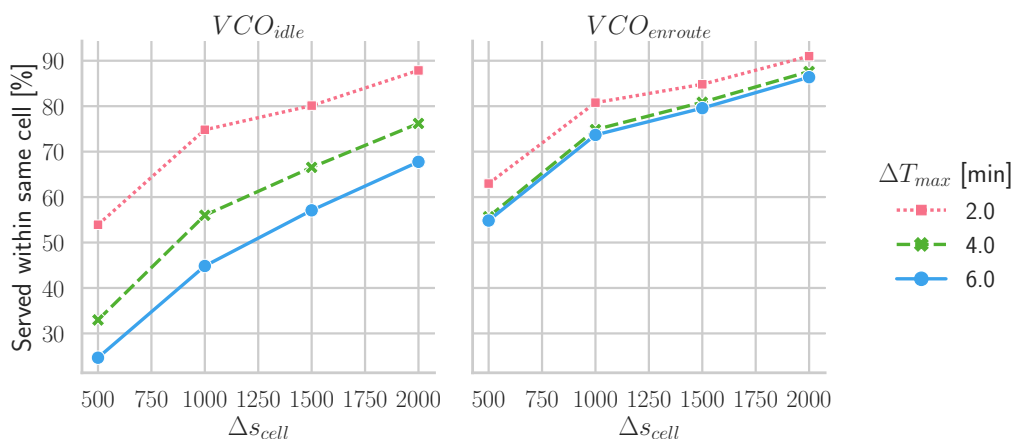
Figure 4.6.: Percentages of customers served by vehicles from the same zone as the customer pickup location. The figures show simulations with ASM travel times.

for Δs_{cell} of 500m, $S_\%^{region}$ increases from 24.7% for ΔT_{max} of 6 minutes to 53.9% for ΔT_{max} of 2 minutes, and for Δs_{cell} of 2km it increases from 67.8% for ΔT_{max} of 6 minutes to 87.7% for ΔT_{max} of 2 minutes. For VCO$_{enroute}$, the $S_\%^{region}$ is found to be generally high and not heavily affected by the value of ΔT_{max}. A major reason for this behavior is that the vehicle is assigned from its availability point and the VCO mainly assigns vehicles based on smaller pickup costs.

With the above simulation results, the section derives the following main observations:

- As the network travel time increases, more customers are served from within the same region as the customer pickup location. This number rises further as the region size increases. However, a significant proportion of customers are still served from other regions.

- A lower ΔT_{max} causes a higher number of customers to be served within the same region. The effect is more apparent for VCO$_{idle}$.

In view of the above observations, the section concludes that the strong assumption of independent regions may hinder the full potential of any FC method used for positioning the AMoD vehicles, especially when the regions are small and contiguous. Such an assumption would also misrepresent the supply-demand imbalance used by any spatial metric, as the supply-demand imbalance is not only dependent on individual regions rather on the neighboring regions as well. Thus, in contrast to the independent regions, the next section analysis the effectiveness of ID method for defining the service quality.

4.3.2. Analysis of Imbalance Density

The previous section showed that the simulated vehicle assignments using VCO$_{idle}$ and VCO$_{enroute}$ are not strictly bound to individual regions. As such, spatial metrics for measuring service quality should also consider the imbalance of the neighboring areas. The ID based spatial formulation resolves this problem. Thus, this section studies how well the ID function describes the AMoD service quality for the Manhattan case study.

As described before, the locations of the customers and the vehicle availability points in each VCO batch are saved during the simulation. This is later used to calculate the ID function for each VCO batch. Let $\rho_{imb}^b(x)$ represent the ID function calculated for a particular VCO batch b. The
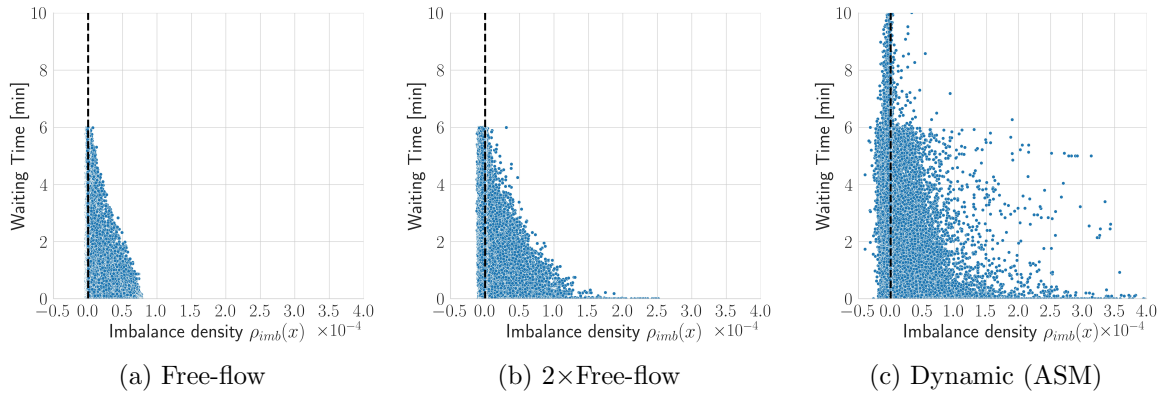
(a) Free-flow

(b) 2×Free-flow

(c) Dynamic (ASM)

Figure 4.7.: The pickup waiting times of the served customers against the values of $\rho_{imb}(x)$ for VCO_{idle}, 3000 vehicles, ΔT_{max} of 6 minutes. For clear visualization, the dotted black line marks the place where x-axis is zero.

section aims to study if these ID functions could serve as a good spatial metric to measure AMoD service quality offered to each customer in b. This is done by evaluating the values of $\rho^b_{imb}(x)$ at the customer pickup locations in b. These ID values at the pickup locations are then used to see if they could describe or differentiate the service quality offered to the customers located at those points; if it does, then it shows that ID could be used as an effective spatial metric to measure service quality in AMoD services, and thus, the associated heat-map formed by the ID relation provides a measure of the actual service quality offered in the area of operation. The section mainly focuses on two important service quality indicators: waiting time to pick up and the probability of being served by an AMoD vehicle.

Relation with Customer Waiting Time

Following the above mentioned procedure, Figure 4.7 shows the scatter plot of the customer waiting times against the corresponding values of ρ_{imb} at the pickup locations. Note that Figure 4.7 only shows the values for served customers as the unserved customers do not have a waiting time. The first thing to note is the decreasing range of customer waiting times with the increasing values of ρ_{imb}; the customer locations having high ρ_{imb} have shorter pickup distances from available vehicles, and thus, have a lower range of possible waiting times. This also shows that ρ_{imb} accurately combines the effects of individual triangular RFs (the highest reaching probability and lowest waiting time at the center and a linear decrease based on euclidean distance) into a single formulation and still distinguishes the waiting time range of individual customers. This observation is more pronounced for the simulations with constant travel times, i.e., Figure 4.7a and Figure 4.7b. In contrast, the ASM method shows anomalies in the relation of ρ_{imb} and the customer waiting times. The main reason is that the ASM method updates the travel times every 30 minutes without canceling the original vehicle assignments even if this causes a delay to the original plan. Therefore, some customers can have a different pickup time than the one calculated while solving VCO. Since ρ_{imb} is computed using the network state at the time of solving VCO, the waiting times of these specific customers show anomalous behavior of having large waiting times.

To study the above relationship further, Figure 4.8 shows the mean and 90th percentile of ρ_{imb} when aggregated using bins of size 5×10^{-6}. The mean and 90th percentile show a decrease in customer waiting times with the increase in ρ_{imb}. The anomalous behavior of ASM method due to period travel time updates can also be observed in Figure 4.8. Nevertheless, the relation of ρ_{imb} and
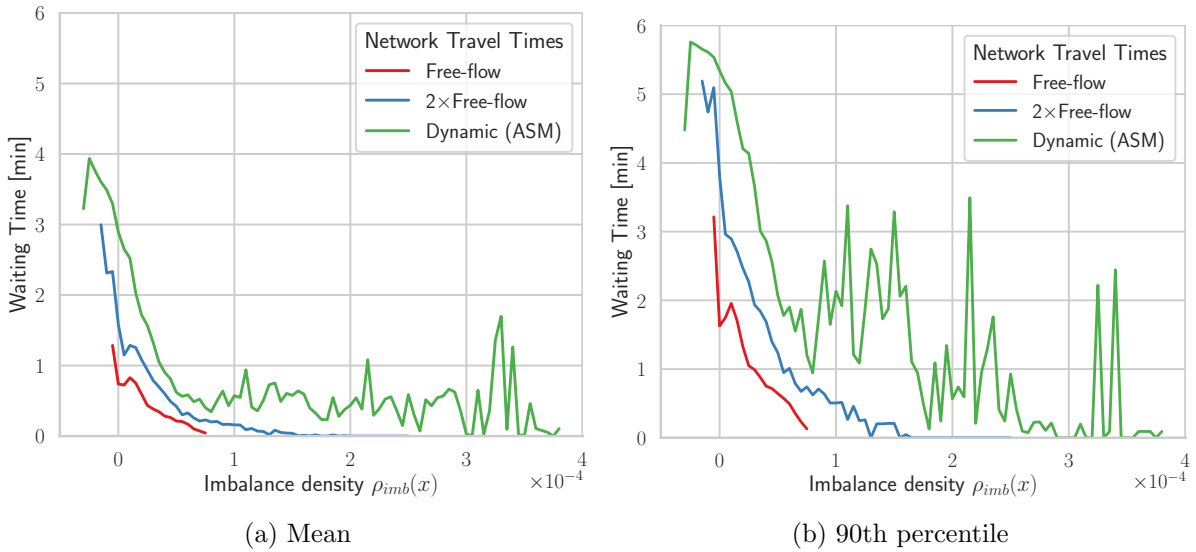
(a) Mean

(b) 90th percentile

Figure 4.8.: The aggregated values of $\rho_{imb}(x)$ using bins of size 5×10^{-6} against the pickup waiting times of the served customers. The simulations used VCO_{idle}, 3000 vehicles and 6 minutes for ΔT_{max}.

waiting times can still be observed for the ASM method: the higher the values of ρ_{imb} the lower the customer waiting times.

Another important observation is that some served customers have a negative ρ_{imb}. This is not counter-intuitive to the used method; a negative value of ρ_{imb} does not imply that the customer will not be served by any vehicle; rather, it only indicates a deficiency of vehicle supply compared to vehicle demand. Consequently, for the customers within areas of negative ρ_{imb}, the underlying VCO must choose which customers should be assigned to vehicles — all of them cannot be assigned simultaneously due to lack of vehicle supply. It is also possible for some of these customers to be located very close to an idle vehicle and still have a negative ρ_{imb}. This is because ρ_{imb} considers the impact of multiple customers and vehicles within reachable distances from the location. Thus, the waiting time of customers with negative ρ_{imb} can range from zero to ΔT_{max} depending on the decisions of VCO.

In contrast, as shown in Figure 4.8b, a sharp reduction in the waiting time is observed as soon as the ρ_{imb} of the served customers shifts from a negative to positive value — a shift from customers located in the areas of vehicle undersupply to areas where vehicles are in abundance. This means that the VCO can assign the nearest vehicle to these customers without competition with other customers due to vehicle oversupply, resulting in a sharp decrease in waiting times. A higher magnitude of positive ρ_{imb} would imply that the associated customers have access to an even higher number of available vehicles closer to the customer locations. This means that the VCO will have multiple options of vehicles to assign to a single customer, out of which it will assign the nearest vehicle. This further reduces the customer waiting time for higher ρ_{imb}.

Figure 4.8 also shows that the maximum and minimum value of the overall ρ_{imb} is dependent on network travel times. A slower city network will have many locations with smaller outreach (smaller RF bandwidths h_i). Thus, the regions with these locations will have a concentrated or denser "reachability" due to smaller h_i — the magnitude of ρ_{imb} will be higher, but it will spread on a smaller area. This is also visible from Figure 4.8, as the network speed becomes slower, the maximum and minimum possible values of ρ_{imb} increase.

Finally, the relationship of ρ_{imb} and waiting times also heavily depends on if the considered AMoD vehicles in the ρ_{imb} are idle or not. The main reason is that ρ_{imb} depends heavily on the RF bandwidths used while calculating the supply density (Eq. 4.5), which assumes that the vehicles can immediately leave from its current location to serve the customers. However, for the enroute vehicles, even though they will become idle at the availability point before the end of ΔT_{max}, they will not be available for the full duration of ΔT_{max}. Their reachable distances will differ from currently idle vehicles (i.e., based on complete ΔT_{max}). Therefore, for the enroute vehicle, ρ_{imb} will wrongfully assume larger reachable distances based on ΔT_{max} while the vehicle will be only partially available within this period. Secondly, the RF used in the dissertation (2D linear kernel) assigns the highest serving probability value at the center of the RF. This assumption directly correlates with the customer waiting times for idle vehicles, as they can immediately serve any customer located at their availability point. But for the enroute vehicles, this is not the case; they can only serve a customer at their availability point after they have completed their assigned journey. Thus, the waiting times only strongly correlate with the idle vehicles for $\text{VCO}_{enroute}$. Since $\text{VCO}_{enroute}$ assigns a big portion of customers to such enroute vehicles, the relationship described in this section between ρ_{imb} and the customer waiting time does not apply to $\text{VCO}_{enroute}$. On the contrary, as all the considered vehicles in VCO_{idle} are idle, the relationship strongly applies to VCO_{idle}.

Relation with Probability of being Served by AMoD service

Besides the customer waiting times, it is also important to study how well ρ_{imb} could describe which customers are served by an AMoD service. Figure 4.9 shows the distribution of ρ_{imb} value for the served and unserved customers along with the percentage of served customers $(S_{\%}^{\rho})$ within different value ranges of ρ_{imb}. As explained in the previous section, a negative value of ρ_{imb} would imply a deficiency in the supply of vehicles in the neighborhood of the customers. In these cases, the VCO must choose which customers are assigned to vehicles due to a lack of sufficient supply of vehicles to serve the customers within ΔT_{max}. This phenomenon is also visible in Figure 4.9: the $S_{\%}^{\rho}$ is significantly small for the negative value of $S_{\%}^{\rho}$. Interestingly, even though the VCO_{idle} serves most of the customers (98.1%) for the free-flow travel time, the ρ_{imb} is still able to delineate this small percentage of unserved customers into a negative range of ρ_{imb}.

As the value of ρ_{imb} increases, the $S_{\%}^{\rho}$ also increases. A sharp increase in $S_{\%}^{\rho}$ happens as the value of ρ_{imb} shifts from negative to positive value. Thus, most unserved customers lie in the negative ranges of ρ_{imb}. This phenomenon is even more visible for the VCO_{idle} as it accurately captures the reachable distances due to consideration of idle vehicles only, as discussed in the previous section. A similar drastic increase in $S_{\%}^{\rho}$ is also visible for $\text{VCO}_{enroute}$ to a lesser extent than VCO_{idle}; the $\text{VCO}_{enroute}$ takes into account the enroute vehicles in ρ_{imb} which may not have the same reachable distances as assumed by RFs. Therefore, some of the unserved customers also lie in the positive value ranges of ρ_{imb} for $\text{VCO}_{enroute}$, which ρ_{imb} marked as having access to a higher number of available vehicles, but in reality, the enroute vehicles were only available for a portion of ΔT_{max}. Nevertheless, the results show that ρ_{imb} is still able to demarcate the unserved customers for $\text{VCO}_{enroute}$, whom it places at lower magnitudes of ρ_{imb}.

The relationship of ρ_{imb} with the probability of being served by the AMoD service for $\text{VCO}_{enroute}$ is in complete contrast to its relation with the waiting time for $\text{VCO}_{enroute}$; the customers with higher values of ρ_{imb} does not exhibit any relationship to the customer waiting times for $\text{VCO}_{enroute}$. This is because even if the enroute vehicles are not immediately available in $\text{VCO}_{enroute}$, they can still reach the destination before the end of ΔT_{max} and be part of VCO. Since VCO tries to assign as many customers as possible, it will still assign customers to enroute vehicles according to their availability time and ΔT_{max}. Thus, a higher value of ρ_{imb} still indicates an increased number of

(a) Free-flow travel time

(b) 2×Free-flow travel time
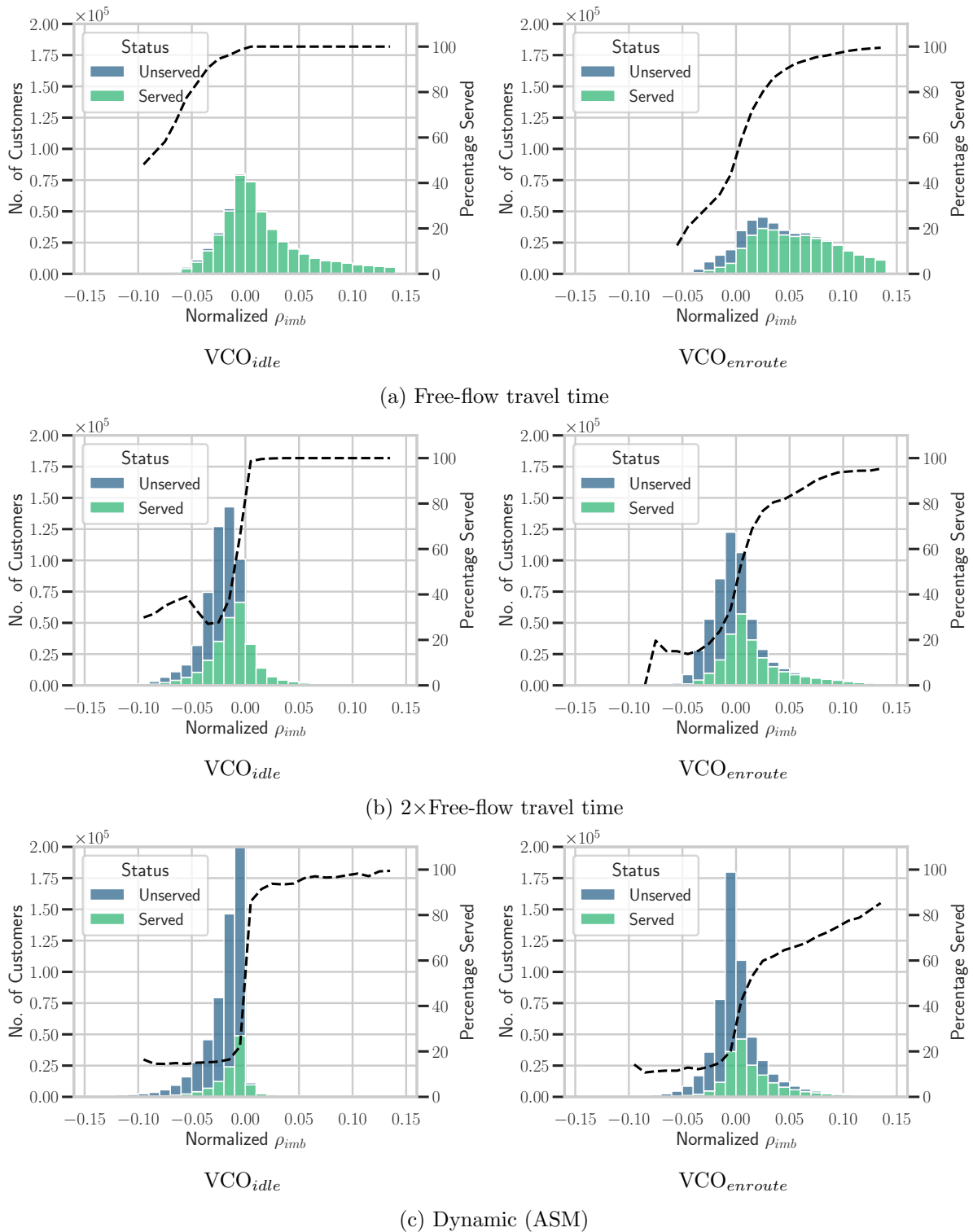
(c) Dynamic (ASM)

Figure 4.9.: The distribution of imbalance density values at the pickup locations of customers for all VCO batches. The values were first normalized using the maximum value of each VCO batch to cover the whole range of the distribution.

vehicles that could be assigned to a customer; however, not necessarily with shorter waiting times.

Imbalance Density (ρ_{imb}) as a Spacial Metric

The previous two sections showed that an ID based relation (ρ_{imb}) can accurately describe the AMoD service quality. It shows a strong relationship with the VCO decisions that which customers are served by AMoD vehicles. In this regard, the VCO_{idle} showed a drastic increase in the percentage of served customers as the value of ρ_{imb} changes from negative to positive. This increase was slightly less drastic for $\text{VCO}_{enroute}$. In general, the results showed that the higher value of ρ_{imb} is for a group of customers, the higher the chances are that they will be assigned to AMoD vehicles. Additionally, VCO_{idle} showed a similar relation of ρ_{imb} with the customer waiting time.

The above outcomes allow the ρ_{imb} to be used as an important metric to measure and visualize the service quality offered by an AMoD service as a heat map. Figure 4.10 shows examples of such a heat map for VCO batches at different times of the day along with the traditional approach of visualizing the supply-demand imbalance in the form of aggregated values.

The first observation is that the aggregated values of supply-demand imbalance over regions help understand the situations in different parts of the operation area. Still, they do not link other regions together in terms of their reachability. In contrast, the ρ_{imb} is not bound by regions to describe the service quality. It exactly demarcates the areas of potential vehicle undersupply without defining explicit regions. Since the shift from negative to positive values of ρ_{imb} marks a drastic increase in the percentage of served customers, a boundary could be drawn to visualize areas of potential undersupply and oversupply of vehicles. Furthermore, a higher negative value of ρ_{imb} can help to identify regions of severe undersupply, as shown for the 12 pm and 6 pm cases in Figure 4.10, where a concentration of unassigned customers can be found near the center.

It is also important to note that since ρ_{imb} represents reachability *density* and not the absolute value of the supply-demand imbalance. Therefore, besides the magnitude of ρ_{imb}, it is also important to consider how big of an area the undersupply of vehicles is spread. For example, for the 6 am scenario in Figure 4.10, even though the customers were located in the negative ρ_{imb} area, all of them could be assigned to vehicles. The main reason is that both the magnitude and the spread of the negative ρ_{imb} were insignificant, and the VCO could assign vehicles to all of them. This also shows the benefit of using ρ_{imb} over aggregated values of supply-demand; most of these customers were located in regions with undersupply of vehicles and, therefore, according to the assumption of independent regions, should not have been assigned vehicles. However, the vehicles from the neighboring areas could serve all the customers.

The usage of ρ_{imb} also shows its benefits over other regional approaches that assume that the vehicles can always serve the customers from the nearby regions without taking into account the inter-regional travel times, like the hierarchical approach [Fagnant and Kockelman, 2014]. For example, for the 12 pm and 6 pm scenarios in Figure 4.10, multiple vehicles were available in the southeast of Manhattan. Yet, some customers remained unassigned in the nearby cells. The hierarchical approach would assume that the vehicles from nearby regions would serve these. However, ρ_{imb} shows that these customers are in the negative part of ρ_{imb}; even though the vehicles may appear geographically closer to these customers, the consideration of actual travel times in ρ_{imb} highlighted that these vehicles may not be able to serve these customers. Another similar example can be observed for the extreme south end of Manhattan for the 6 pm case, where a tiny area of vehicle undersupply is formed, and only one customer remained unassigned in the area.

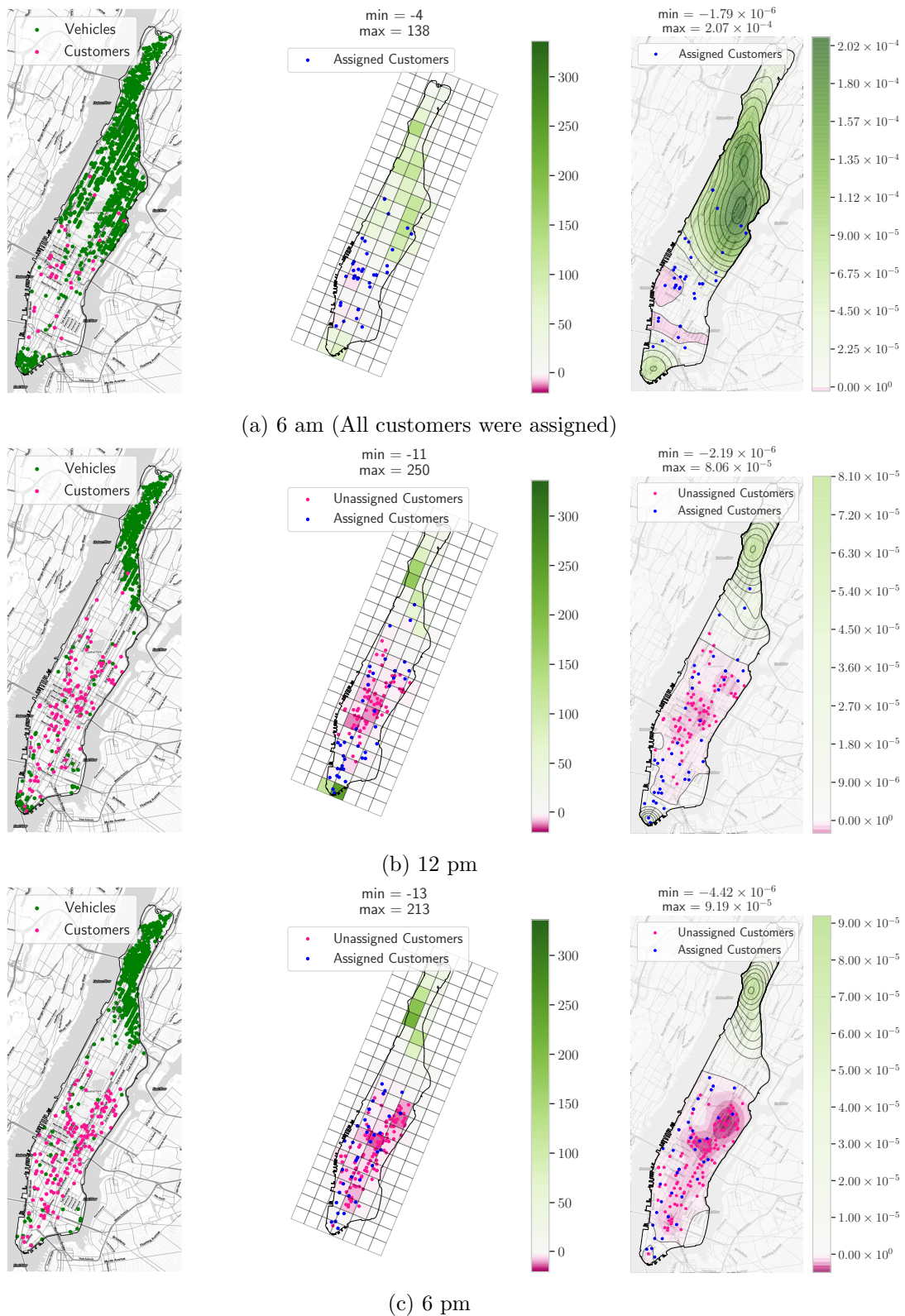(a) 6 am (All customers were assigned)

(b) 12 pm

(c) 6 pm

Figure 4.10.: The ID function as spatial metric to visualize the AMoD service quality in the form of a heat map. The color bar indicates the supply-demand imbalance with pink and green marking the vehicle undersupply and oversupply, respectively. The simulations in the figure are with VCO_{idle}, 3000 AVs, and ASM travel times.

4.4. Conclusion

The AMoD service area is often divided into a disjoint set of regions for analytical, operational, or privacy reasons. This dissection is usually based on the assumption that these regions are independent, which means that an AMoD vehicle located in a specific region only serves the customers in that region. However, depending on the region sizes and the network state, an AMoD vehicle is not bounded by such restrictions. Therefore, this chapter introduced a density-based spatial metric, called ID, that describes the AMoD service quality offered in the operation area without relying on the definition of regions. It links the individual locations of customers and vehicles with the overall AMoD service quality in the form of a heat map. It uses 2D linear functions as RFs to model the reachable distances and the probability of reaching the surrounding area from the centers of RFs. The chapter studied the effectiveness of the ID relations to describe the service quality when VCO_{idle} and $VCO_{enroute}$ are used for assigning vehicles to customers.

The ID showed strong relation to the assignment decisions of VCO_{idle} and $VCO_{enroute}$. It describes the customers who are probably assigned an AMoD vehicle. The ID also showed a strong relationship with the customer waiting times for VCO_{idle}. From these observations, the chapter concludes that ID provides an important metric for measuring the AMoD service quality offered in the operation area. The chapter did not focus on using ID relations for actively improving the AMoD service quality; instead, the main focus was on proving its effectiveness as an essential spatial metric. The following chapters study how ID could be used to improve the AMoD service quality.

Chapter 5.

Reachability Functions based Repositioning Strategies

The previous chapter introduced a novel spatial metric, the imbalance density (ID), to measure the service quality in AMoD services. It was based on the concept of reachability function (RF) which consistently takes into account the proximity of neighboring locations. This helps to describe the AMoD service quality offered to individual customers in the form of a heat map. This chapter further extends the concept of ID relations to improve the AMoD service quality actively. In this regard, the main focus is on repositioning the AMoD fleet to minimize the supply-demand imbalance caused by the disparity between customer origin and destination locations. Thus, the chapter presents a repositioning approach that utilizes the ID relations to reposition idle vehicles.

The chapter is structured as follows. Section 5.1 introduces the primary motivation for developing a new repositioning method. Then section 5.2 discusses the problem formulation used for repositioning along with the forecast method used in the dissertation for estimating the future AMoD system state. Section 5.3 develops a min-distance repositioning method that is not based on ID but can still achieve a significant improvement over the benchmark method used in the dissertation. Then section 5.4 describes the optimization formulation used for ID based repositioning methods and two methods used in the dissertation for its solution. To understand the overall working of these newly developed repositioning strategies, section 5.5 conducts tests on static repositioning instances to analyze the behavior of these methods. Finally, section 5.6 performs a detailed sensitivity analysis of the developed method in the agent-based simulation to evaluate the overall performance improvements.

5.1. Motivation

In the current MoD and Taxi services, the overall number of customers and the locations of their origins and destinations change significantly across and within days. These temporal changes in the pattern of customer requests are expected to be similar in the future for an AMoD service. As mentioned in section 3.2.2, the pattern usually differs significantly between a business day and a weekend. On weekends, leisure activities significantly impact the destination and origin of trips. Contrarily, during a workday, many trips are regular trips from home to work (morning trips) and back home (evening trips). Therefore, depending on the time of the day, there could be an imbalance in the number of trips going to a region and starting from it. A similar phenomenon is observed in the NYC data set. As shown in Figure 5.1a, the imbalance between trip origins and destinations differs significantly for different hours of the day. In the morning, the regional imbalance is shifted towards the center; more people go from residential to commercial areas at the center than the other way around. This is indicated by the concentration of highly positive values around the center at 5 and 8 am in Figure 5.1a. Contrarily, during the evening and night hours, the imbalance shifts towards the residential areas in Manhattan's north, northeast, and northwest.
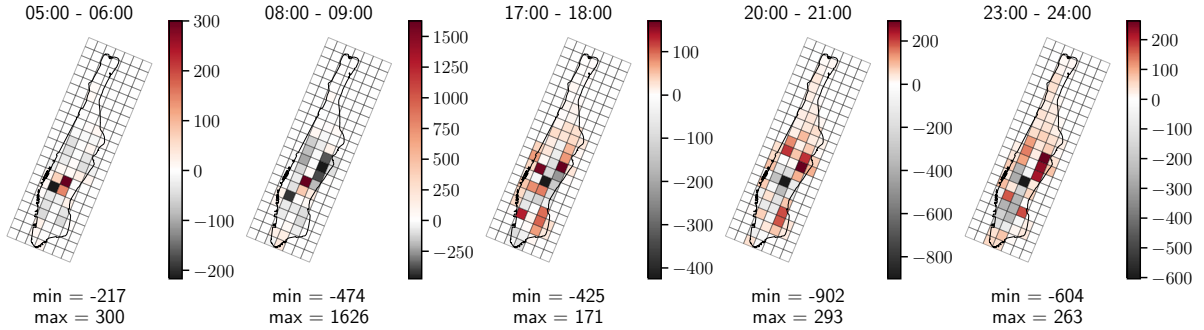
(a) The hourly differences between the number of trips arriving into the regions (destinations) and the number of trips leaving the region (origins) in the original data.
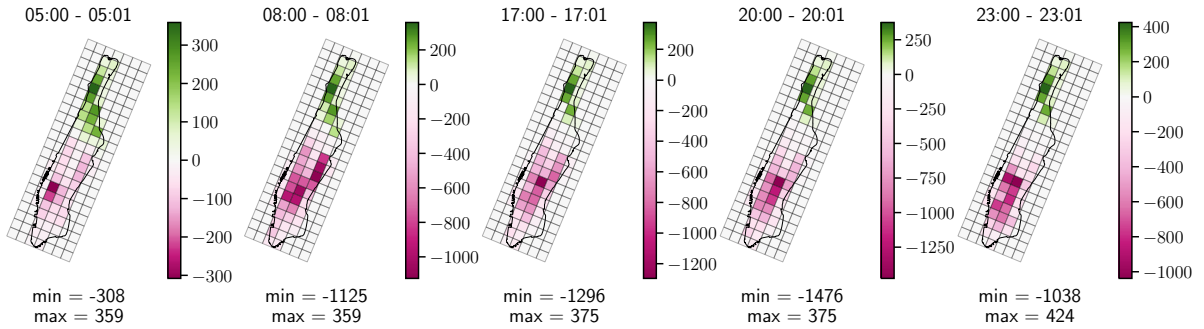
(b) Regional imbalances in the simulated AMoD service without repositioning of idle vehicles
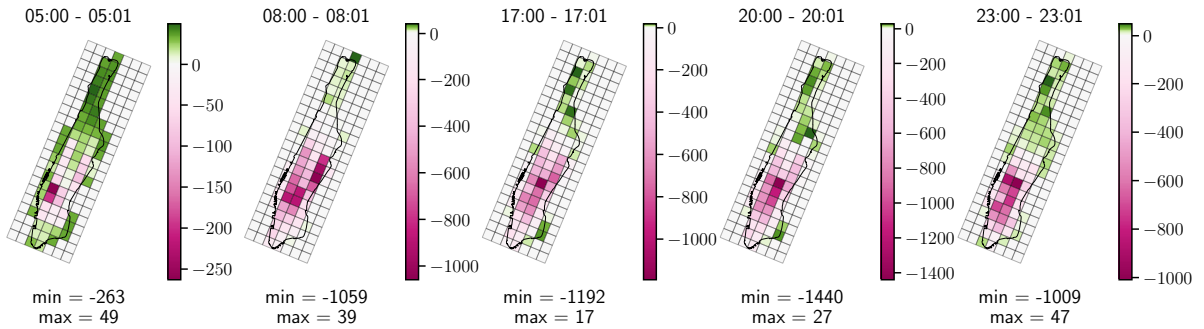
(c) Regional imbalances in the simulated AMoD service with Pavone's method for repositioning of idle vehicles. The simulation used 30 minutes as repositioning and forecast period.

Figure 5.1.: Impacts of repositioning approach for AMoD services simulated using data 6 and 7 June 2016. The time represents the second day of the simulation. The simulations used ΔT_{max} of 6 minutes, 3000 vehicles, $VCO_{enroute}$, ASM travel times and Δs_{cell} of 1 km. Figures (b) and (c) take into account all the available vehicle locations (idle and the enroute vehicles about to finish the trip). The regional imbalance is calculated by subtracting the number of forecast trip origins (within 1 minute period) from vehicle locations.

Due to the above imbalances, a Taxi or an AMoD vehicle dropping off a customer at night may end up in a region where the waiting time for getting the next customer is significantly high. In this case, a human taxi driver would relocate to higher-demand areas based on past experiences. In contrast, a centrally controlled fleet of AVs would require an explicit relocation command from the central FC. Without such a relocation, many AMoD vehicles may remain idle for a long time while a large number of customers in other parts of the city would simultaneously remain unserved. This leads to a significant under-utilization of the fleet.

The above phenomenon is also shown in Figure 5.1b for a simulation with $VCO_{enroute}$ and a fleet of 3000 AVs (the rest of the parameters are the same as in Table 4.1). Since the dissertation simulates two days (6 and 7 June 2016), the figures compare the values on the second day of simulation for a fair demonstration of the phenomenon. This ensures that the observation is least affected by the initial locations of the vehicles at the beginning of the simulation. Figure 5.1b shows an instance of the supply-demand imbalance using the availability locations of vehicles and customer origins aggregated over one minute. The imbalances of destinations and origins lead to the accumulation of vehicles near residential areas at night. Many of these vehicles do not find enough trips to return to the demand-intensive regions, making some of these residential regions have a vehicle oversupply while others have a significant vehicle undersupply. This is visible from the high supply-demand imbalance values in Figure 5.1b. Additionally, since a purely profit-oriented VCO prefers longer trips to earn higher distance-based fare, even during the daytime, customers going away from demand-intensive regions in the center to the far-off areas are preferred, which further maintains an oversupply of vehicles in the far-off regions. This is also indicated by a constant, highly positive supply-demand imbalance in the north of Manhattan. As shown in Figure 4.4 for the ASM travel times, this leads to lower fleet utilization with time.

The commonly used approach to resolve the above supply-demand imbalance is to periodically reposition idle vehicles to demand-intensive regions, referred to as *mid-term* repositioning (section 2.5.4). This approach solves an optimization problem to determine the repositioning of idle vehicles using the demand forecasts of each region. Many studies based their mid-term repositioning approaches on the method of Pavone et al. [2012] where the excess vehicles are equally distributed among regions while minimizing the total VKT [R. Zhang and Pavone, 2016; Ruch et al., 2018; S. Hörl et al., 2019]. Therefore, the current chapter uses the method of Pavone et al. [2012] as the benchmark. The dissertation also tested other methods mentioned in the literature, however, the method of Pavone et al. [2012] outperformed other approaches in the simulation framework. A detailed description of Pavone's method and other benchmark methods is included in the appendix.

Figure 5.1c shows the impact of using Pavone's method. The method leads to a better distribution of fleet with a significant reduction in the accumulation effect of the AMoD vehicles. Pavone's method can also be integrated inside VCO, referred to as *short-term* repositioning [Dandl, M. Hyland, et al., 2020; Dandl, M. Hyland, et al., 2019]. However, the chapter focuses on the improvement of mid-term repositioning as this gives explicit control over the frequency of repositioning and is more practical for actual AMoD services than a frequent repositioning in each VCO batch.

Regardless if the repositioning is done inside VCO or in a separate mid-term optimization, most repositioning formulations do not consider the size and proximity of other regions to determine the system imbalance. This leads to two major issues:

- Region-based repositioning algorithms try to balance as many regions as possible with minimum extra VKT while ignoring the fact that vehicles from the neighboring regions can also be assigned to the customers. Thus, even though Pavone's method distributes the excess vehicles equally among regions, in reality, many of these repositioned or already available vehicles would be assigned to customers from other regions. Thus, the repositioning decisions taken may not

reflect the VCO decisions. This would cause some regions to remain imbalanced while others to have a high number of excess vehicles. Figure 5.1c shows this phenomenon; the supply-demand imbalance is significantly lower than Figure 5.1b (without repositioning), still many regions have severe undersupply while some regions have a significant excess vehicle. This is because the method equally distributes the excess vehicles, which does not send enough vehicles to certain regions to fulfill the local requirements. For example, the equal distribution of excess vehicles sent fewer vehicles to the center which did not fully meet the local demand. This caused a vehicle undersupply in the center. Furthermore, since the vehicles from the neighboring regions can also be assigned to the customers, eventually it led to an even greater area having a vehicle undersupply. On the contrary, the method sent more vehicles than required to the low-demand areas (mainly the north part of Manhattan), which mostly remained underutilized.

- Since the majority of repositioning approaches prioritize the minimization of repositioning VKT, short repositioning trips are prioritized. This can lead to larger areas remaining highly imbalanced although idle vehicles could have been repositioned thereby making longer repositioning trips, such as Upper West Side, Upper East Side, and mid-town in Figure 5.1c.

The current chapter focuses on resolving the above issues by introducing RFs based repositioning methods. Ultimately, the goal is to position the idle vehicles in such a way that the pickup points of the potential customers remain in the areas of positive ID in each VCO batch. If the values of ID are increasingly positive at the customer pickup locations, the chances of AMoD fleet successfully serving these customers are equivalently higher.

5.2. Problem Formulation

This section formulates the repositioning problem for the AMoD services. It first describes a general repositioning problem that considers all coordinates inside the operational area. Then it describes the region-based repositioning problem that deals with AMoD supply and demand aggregated over regions.

5.2.1. General Repositioning Problem

In an AMoD scenario, the spatiotemporal distributions of customer requests and vehicle availability play a major role in the overall service quality. As described in section 5.1, after dropping customers to their destinations vehicles might take a long time before getting the next customers, despite the fact that many customers remain unserved in other parts of the city. Therefore, a repositioning algorithm should be able to route idle vehicles to areas of higher demand to minimize the supply-demand imbalance. Ideally, such an algorithm should consider all possible coordinates within the city for this purpose. However, an accurate supply and demand forecast is usually not available on such a granular scale. Additionally, developing a repositioning algorithm for such a fine scale drastically increases the computational complexity. Therefore, the dissertation uses a region-based repositioning strategy as discussed in the following section.

5.2.2. System Forecast and Region-Based Repositioning Problem

Typically, regions are introduced to aggregate supply and demand forecasts and to limit the possible repositioning destinations in the optimization formulation. Let Z denote a disjoint set of zones in the AMoD operation area. Let t represent the simulation time at which the repositioning decisions are

made. Then the vehicle supply and demand forecasts of each zone are made for a temporal period $[t, t + \Delta T_h]$ where ΔT_h describes the forecast horizon. A repositioning problem is solved periodically after every ΔT_r with weights $\omega_z \in W_Z$ representing the imbalance of supply and demand for each zone $z \in Z$. The repositioning problem aims to redistribute the idle vehicles among regions in a way that minimizes the regional supply-demand imbalances ω_z. The dissertation breaks the ω_z into two parts: the regional estimate of vehicle supply ω_z^{sup} and the regional estimate of vehicle demand ω_z^{dem} such that $\omega_z = \omega_z^{sup} - \omega_z^{dem}$.

For the efficient performance of any repositioning method, the accuracy of the forecast information is extremely important. It requires a forecast of future AMoD system state, mainly consisting of the future vehicle availability and expected customer demand. Even if the customer demand is known accurately, predicting the future vehicle supply is not always straightforward; it heavily depends on the VCO decisions. For example, after the repositioning decisions are made, the VCO may unexpectedly assign a high number of vehicles to customers going in the same direction as the repositioning vehicles, leading to a different system state than expected at time t. Ideally, the used forecast method should also consider the possible long-term VCO decisions. However, investigating such an accurate forecast method is complex and beyond the scope of the current dissertation. Therefore, the dissertation uses a simpler approach to estimate the future zone imbalances W_Z as described below.

Fundamentally, the dissertation counts the number of vehicles that will be available after dropping off a customer inside a region z as positive and subtract the number of trips starting from z. The dissertation uses the sets R_z^+ and R_z^- representing the forecast of customer drop-off and pickup locations for zone z, respectively. It should be noted that R_z^+ and R_z^- can only be roughly estimated since the exact number of vehicles that will end up in z after dropping off customers within $[t, t + \Delta T_h]$ is not known at time t. Therefore, the dissertation calculates them using the customer data used: the customer destinations and origins within time horizon $[t, t + \Delta T_h]$ are summed up to calculate R_z^+ and R_z^- regardless if they are actually served by AMoD service or not.

Following the above strategy, two types of demand forecasts are possible depending on the value of ΔT_h used. If ΔT_h is a positive value, the dissertation uses the information of the future customers who are not yet revealed to the system, referred to as *perfect forecast*. In reality, such an accuracy is not available to the operator. Therefore, the dissertation also tests forecasts with lower accuracy by using a negative value for ΔT_h. In this case, the dissertation aggregates the information of recent customers within the last ΔT_h period, referred to as *imperfect forecast*. To estimate the W_Z, the dissertation additionally uses the values V_z^+ representing the idle vehicles in zone z, and V_z^r representing the vehicles currently enroute to zone z for repositioning from the previous cycle. Since the dissertation uses regular grids with cell size Δs_{cell} for dividing the operation area into regions, the accuracy of the forecast also depends on Δs_{cell}; the spatial accuracy of the forecast is inversely proportional to the cell size value Δs_{cell}.

Using the above notations, the dissertation uses the R_z^- as the demand side estimate ω_z^{dem}. For the supply side estimate ω_z^{sup}, the dissertation uses two variants. The first variant includes the estimated vehicle arrivals using customer data $|R_z^+|$. This assumes that the customer destination points may provide an estimate of vehicle supply as the AMoD vehicles assigned to these customers will end up in the corresponding destination zone. Thus, the formulation is given as:

$$\omega_z^{sup(c)} = |V_z^+| + |V_z^r| + |R_z^+| \tag{5.1}$$

where the superscript (c) indicates that the customer data is used in the estimation of vehicle supply. However, since R_z^+ may not necessarily represent the vehicle availability as many of the customers may not be served (the vehicle may not drive with the customer to z), the dissertation also examines
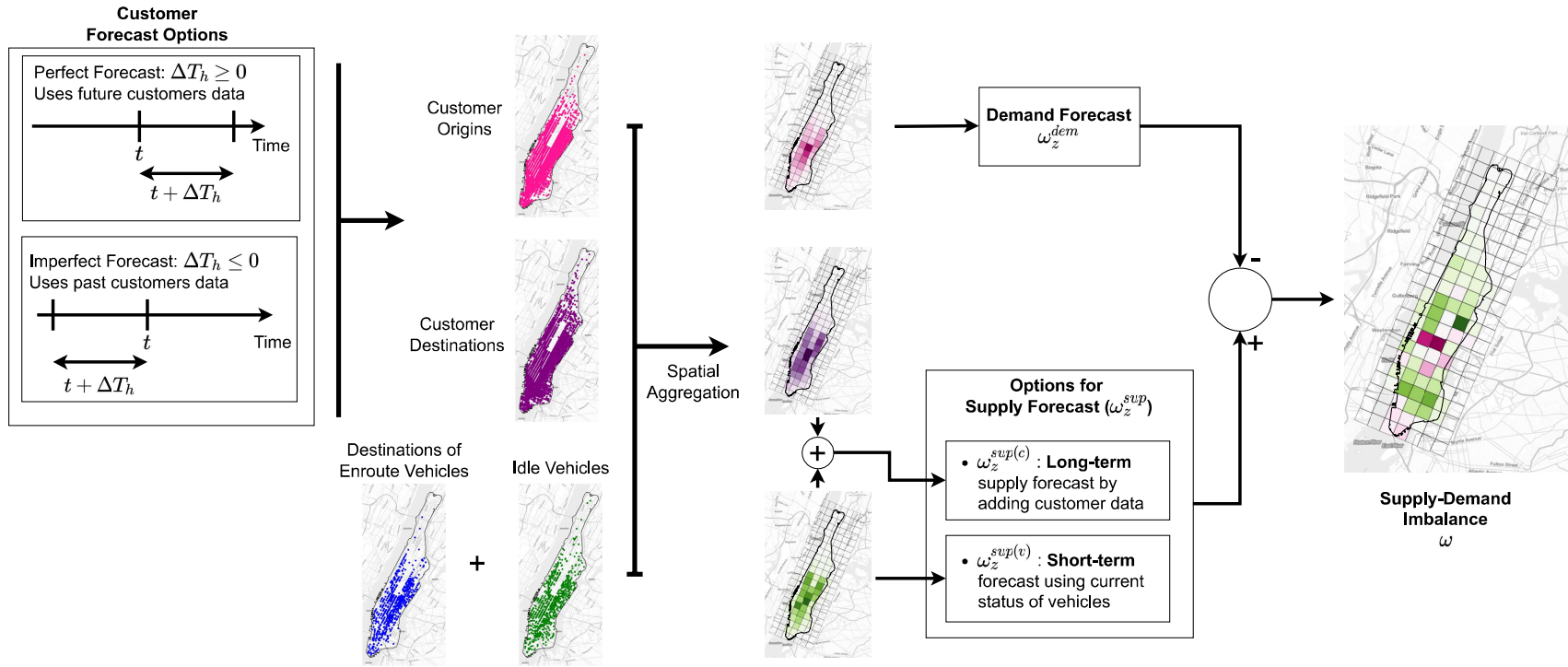
Figure 5.2.: The calculation of regional supply-demand imbalance used in the dissertation.

the formulation where it is removed from supply forecast, given as:

$$\omega_z^{sup(v)} = |V_z^+| + |V_z^r| \tag{5.2}$$

where the superscript (v) indicates that only current vehicle data is used in the estimation of supply. In the following section, wherever the dissertation uses ω_z^{sup} it means that either of the above equations can be used to calculate ω_z^{sup}. Figure 5.2 summarizes the calculation of the regional weights for the supply-demand imbalance.

Let surplus zones $Z^+ \subseteq Z$ and deficiency zones $Z^- \subseteq Z$ represent the zones with positive and negative weights ω_z, respectively. Then the repositioning algorithm called at time t returns a flow matrix $\boldsymbol{u}^{(t)} \in \mathcal{I}^{|Z^+| \times |Z^-|}$ that represents the number vehicles that need to be repositioned from vehicle surplus to vehicle deficiency zones. Note that the repositioning algorithms are allowed to only reposition idle vehicles V_z^+ from zone z. Typically, the repositioning problem also considers a cost for each repositioning decision in $\boldsymbol{u}^{(t)}$, given by the cost matrix $\boldsymbol{C}^{(t)} := (c_{ij}^{(t)}) \in \mathbb{R}_{\geq 0}^{|Z^+| \times |Z^-|}$. It can be provided in terms of traveling cost, traveling time, or the marginal profit expected by serving additional customers after repositioning. The dissertation uses the $\boldsymbol{C}^{(t)}$ in terms of inter-regional travel distances.

It should also be noted that ω_z represents the expected vehicle imbalance in zone z within the time horizon ΔT_h in case the operator does not reposition. Most region-based repositioning algorithms assume that a single repositioning vehicle changes the imbalance weight by one, and thus, the post state of the region after repositioning is estimated as:

$$\omega_z + \sum_{i \in Z} u_{iz} - \sum_{j \in Z} u_{zj} \tag{5.3}$$

This is a simplification as the repositioning vehicle will only be available for a part of the time horizon ΔT_h in the destination zone. In fact, if the distance between regions i and z is too large to be reached within ΔT_h, it is even possible that the repositioning vehicle is not present at all during ΔT_h.

The explicit repositioning decisions, i.e., send vehicle v to position x, have to be derived from the region-based flow matrix $\boldsymbol{u}^{(t)}$. There are several ways to do that. The dissertation first creates the set of all trips originating from zone z. Then, a greedy algorithm randomly picks the next destination zone and sends the idle vehicle in z that has the shortest distance to the centroid of the destination zone. This process is repeated till all vehicles are assigned to repositioning points. All vehicles are repositioned to the node closest to the centroid of the destination zone.

Henceforth, the superscript in flow matrix $\boldsymbol{u}^{(t)}$ and $\boldsymbol{C}^{(t)}$ are dropped since it is clearly understood that a particular instance of the repositioning problem at time t is solved in the following sections.

5.3. Min-distance Repositioning Strategy

The first approach tested in the dissertation is referred to as the *min-distance* approach where the aim is to minimize the overall supply-demand imbalance with minimum VKT. In order to achieve this, it tries to reposition as many idle vehicles as possible from the surplus regions to deficiency regions with hard constraints on the number of required vehicles in the destination zone. The method is formulated as:

$$\min_{\boldsymbol{u}} \quad \sum_{i \in Z^+} \sum_{j \in Z^-} c_{ij} u_{ij} - M \sum_{i \in Z^+} \sum_{j \in Z^-} u_{ij} \tag{5.4a}$$

$$\text{s.t.} \quad \sum_{j \in Z^-} u_{ij} \le \min(\omega_i, |V_i^+|) \qquad \qquad \forall i \in Z^+ \tag{5.4b}$$

$$\sum_{i \in Z^+} u_{ij} \le -\omega_i \qquad \qquad \forall j \in Z^- \tag{5.4c}$$

$$Z^+ = \{z | z \in Z, \omega_z > 0\}$$

$$Z^- = \{z | z \in Z, \omega_z < 0\}$$

$$\boldsymbol{u} \in \mathbb{Z}_{\ge 0}^{|Z^+| \times |Z^-|}$$

where M in Eq. 5.4a is a big number to maximize the number of repositioned vehicles while the first term minimizes the traveled VKT. Note that without the second term in the objective function, there will be no repositioning. Eq. 5.4b adds the constraint that the number of repositioned vehicles originating from zone i must be less than the expected excess vehicles in zone i. This ensures that the repositioning method does not reposition excessive vehicles out of a surplus zone i and that enough vehicles are left in the zone to fulfill the local demand. Similarly, Eq. 5.4c says that the number of repositioned vehicles arriving at a zone must be less than or equal to the expected deficiency in the destination zone. This ensures that the destination zones do not receive excess vehicles.

The main difference between Pavone's method and the min-distance method is that while Pavone's method tries to equally distribute the excess vehicles among all regions, min-distance tries to send just enough vehicles to destination zones to fulfill the demand. Thus, unlike Pavone's method, the number of repositioning vehicles received by each zone also depends on the demand in the destination zones.

5.4. Reachability Function based Repositioning Strategy

This section introduces the reachability function based repositioning (RFR) methods. The RFR methods fundamentally try to close the gap between supply and demand densities via the repositioning of idle vehicles. First, a general formulation of the RFR method is introduced in section 5.4.1 where it is assumed that the AMoD operator has the exact knowledge of future customers and vehicle supply. Then, section 5.4.2 and section 5.4.3 present region-based approaches for minimizing the expected ID via repositioning. Finally, section 5.4.4 discusses the two optimization approaches used in the dissertation for solving the new repositioning formulations.

5.4.1. Repositioning using Exact Coordinates

This section presents the general version of RFR method where it is assumed that the MSP has exact knowledge of the origins and destinations of the future customer requests within the forecast time horizon, represented by sets R_- and R_+, respectively. It uses the exact customer and vehicle locations for calculating reachability density (RD) values. Furthermore, the section discusses the involved computational challenges for such formulation and the procedures that can be adopted to reduce computational complexity.

Let $\bar{p}_{R_-}$ and $\bar{p}_{R_+}$ be the densities calculated using Eq. 4.4 for pickup and drop-off locations, respectively. It is assumed that $\bar{p}_{R_-}$ and $\bar{p}_{R_+}$ are calculated using appropriate bandwidths for each

location according to the current state of the city network. The pickup and drop-off locations describe demand and supply because vehicles are required at the pickup locations and become available at drop-off locations[1]. Thus, $\bar{p}_{R_-}$ and $\bar{p}_{R_+}$ indicate the expected demand and supply of vehicles, respectively. Let $x_1^{idle}, x_2^{idle}, ..., x_{n_{idle}}^{idle} \in \mathbb{R}^2$ be the positions of idle vehicles and $x_1^r, x_2^r, ..., x_{n_r}^r \in \mathbb{R}^2$ be the points in the operating area where idle vehicles could be repositioned and become part of the vehicle supply. Note that multiple vehicles can be repositioned to the same point, assuming that the repositioned vehicle will park itself to the nearest free geographical position after reaching the destination. The formulated repositioning problem has multiple conflicting objectives and forms a multi-objective optimization problem (section 2.2.2): the FC aims to reduce the overall supply-demand imbalance which would require additional VKT for repositioning while the excessive VKT must be simultaneously minimized to reduce additional traveling cost. Thus, the multi-objective repositioning problem is given as:

$$\min_{\boldsymbol{u}} f(\boldsymbol{u}), g(\boldsymbol{u}) \tag{5.5a}$$

$$\text{s.t.} \quad \sum_{j=1}^{n_r} u_{ij} \leq 1 \qquad\qquad\qquad \forall i \in \{1, ..., n_{idle}\} \tag{5.5b}$$

$$g(\boldsymbol{u}) = \sum_{i,j} \bar{c}_{ij} u_{ij} \tag{5.5c}$$

$$f(\boldsymbol{u}) = F\big(\underbrace{m(x, \boldsymbol{u}) + \bar{p}_{R_+}}_{\text{supply density}}, \quad \underbrace{\bar{p}_{R_-}}_{\text{demand density}} \big) \tag{5.5d}$$

$$m(x, \boldsymbol{u}) = \underbrace{\frac{3}{\pi} \sum_{i=1}^{n_r} \left[\frac{k(x, x_i^r, h_i)}{h_i^2} \sum_{j=1}^{n_{idle}} u_{ji} \right]}_{\text{density of the repositioned vehicles}} + \underbrace{\frac{3}{\pi} \sum_{i=1}^{n_{idle}} \left[\frac{k(x, x_i^{idle}, h_i)}{h_i^2} \sum_{j=1}^{n_r} (1 - u_{ij}) \right]}_{\text{density of the vehicles at original positions}} \tag{5.5e}$$

where $\boldsymbol{u} := (u_{ij}) \in \{0, 1\}^{n_{idle} \times n_r}$ is a binary variables matrix for flow of idle vehicles from origins to repositioning points and $\bar{C} := (\bar{c}_{ij}) \in \mathbb{R}_{\geq 0}^{n_{idle} \times n_r}$ is the matrix of traveling costs. The function $F : \mathbb{R}^2 \mapsto \mathbb{R}$ is a metric for measuring the deviation between the two densities. Thus, Eq. 5.5d is the objective function for minimizing the difference between the customer demand and vehicle supply by reshaping the idle vehicle density via repositioning. The constraint in Eq. 5.5b forces a maximum of only one destination point to be assigned to each idle vehicle.

The formulation is a multi-objective optimization problem with $n_{idle} \cdot n_r$ quadratic decision variables and n_{idle} constraints. This problem could be a large and difficult problem to solve depending on the number of idle vehicles and repositioning points inside the city. Furthermore, the optimization algorithm will have to repeatedly evaluate Eq. 5.5d which is a computationally expensive task depending on n_{idle} and the coordinate resolution used for calculating the RFs (i.e. the function k). The values $\bar{p}_{R_+}$ and $\bar{p}_{R_-}$ could be calculated before the optimization process, but Eq. 5.5e must be evaluated in each iteration of the solution process. Assume that the RFs in Eq. 5.5e are evaluated and saved for k_{eval} equidistant points in the first iteration (that still consumes a lot of memory), the evaluation time of Eq. 5.5e will be still of $\mathcal{O}(n_r k_{eval} n_{idle})$. Additionally, for a considerable accuracy, k_{eval} has to be a large number, for example, for Manhattan with a distance of 50 m between each points the k_{eval} must be in the order of 10^5, and increases to 10^6 for a distance of 10m. Such a running time for the repositioning problem is not practical, especially for a dynamic AMoD scenario where it must be solved periodically.

---

[1]This assumption is only valid for the AMoD use-case without pooling of requests

Therefore, even if an ideal forecast method provides exact locations of future customers, using exact geographical points for a RFR optimization approach would still not be feasible. Even if only the Manhattan network nodes (similar to section 4.3.2) are used, the problem complexity would still be infeasible. There could be two approaches to reduce the required computational effort:

1. Incrementally calculate the new densities in each iteration of the optimization. This can be done by subtracting the kernel contributions of only those kernels whose associated decision variables are changed, i.e., the inner loop of both terms in Eq. 5.5e. However, this is only beneficial for metaheuristic algorithms that have complete control over evaluating objective functions. For MIP solvers, it would be not easy to use such a method.

2. A more suitable and scalable approach is to use the binning technique that is also used for calculating KDE using fast Fourier transform (FFT) method for large data sets [SILVERMAN, 1982]. The following section discusses this technique in detail.

5.4.2. Reachability Function Based Repositioning with Regions

As the previous section showed, using raw coordinates for the RFR method poses a significant computational challenge. Second, as described in section 5.2.2, the customer forecast is also not available with such accuracy. Therefore, instead of directly using the ID, the dissertation uses its approximation over regions for the purpose of repositioning.

The formulation is motivated by the binning technique used for KDEs in which the data points are sorted and distributed into bins with a weight assigned to each bin. After binning, the KDE is calculated either by using weighted KDE formulation or using Fourier transform. In the latter case, the KDE is calculated by multiplying the Fourier transform of the binned data by the Fourier transform of the kernel and then calculating the inverse Fourier transform [SILVERMAN, 1982]. However, for the Fourier transform approach, all kernel bandwidths are required to be the same. Since the RFs in the dissertation have different bandwidths according to the current network state, the Fourier transform approach can not be used for the purpose of repositioning. Therefore, the dissertation uses weights assigned to the centroids of each zone to estimate the RD of vehicles and customers; for a zone $z \in Z$, the number of individual customers or vehicles in each region is used as the aggregated regional weights $\boldsymbol{w} \in \mathbb{R}^{|Z|}$ and the regional centroids as the data points. Thus, a region-based approximation to the RD is given as:

$$\bar{p}_Z(x) = \frac{3}{\pi} \sum_{z \in Z} w_z \frac{k(x, x_z^Z, h_z)}{h_z^2} \tag{5.6}$$

with $\int \bar{p}_Z(x)dx = \sum_{z \in Z} w_z$. The points $\{x_1^Z, x_2^Z, ..., x_{|Z|}^Z\}$ are the zone centroids for zones Z.

Given the above region-based approximation of RD, Figure 5.3 shows the overall flow of the RFR with regions (RFRR) method. It adopts the formulation of $\bar{p}_Z(x)$ in Eq. 5.6 for the purpose of repositioning. However, instead of using the aggregated weights for vehicle locations and customer origins, it uses the forecast of regional weights introduced in section 5.2.2. Thus, instead of calculating the densities for customers and idle vehicles separately, the imbalances for each zone ω_z are first calculated using ω_z^{dem} and ω_z^{sup}. Then a combined imbalance density (ID) using weighted RD function (Eq. 5.6) is considered.

The formed optimization method has multiple objectives: first, for the expected supply-demand imbalance, and second, for the repositioning VKT. It should be noted that even though the dissertation uses a regular grid to define regions, the repositioning method is also valid for arbitrary regions; the only major requirement of the RFRR approach is that the kernel locations are predefined and
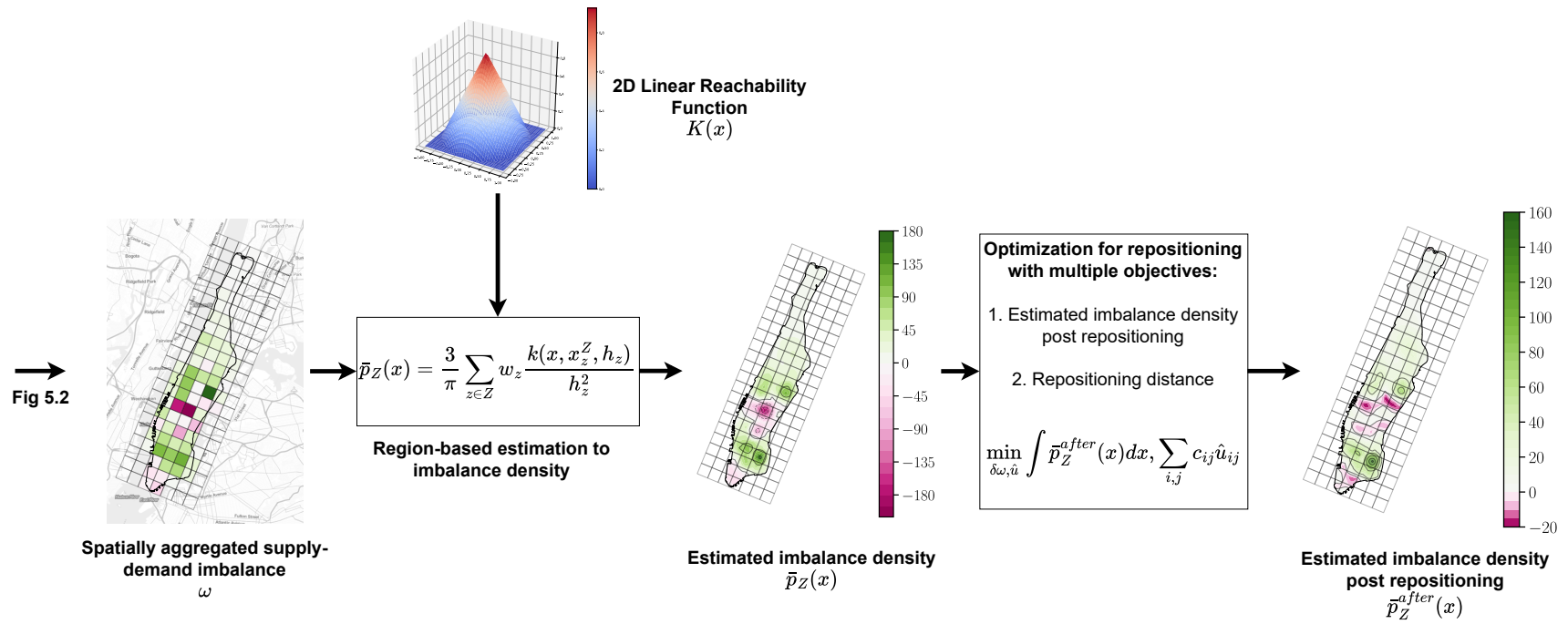
Figure 5.3.: The flow of the RFRR strategy with vehicle supply and demand aggregated over zone centroids.

fixed. In the case of arbitrary regions, the centroids of the arbitrary regions could be used for this purpose. Additionally, this section introduces a general version of the formulation that purely aims at minimizing the supply-demand imbalance density with minimum repositioning VKT without any restriction on regions: contrary to the traditional restriction of sending vehicles only from surplus to deficiency regions, the idle vehicles can reposition to and from any region. Because of the ability to send vehicles to or from any region, the formulation in the current section is referred to as RFR with regions and full flow (RFRRf). The next section will introduce formulations with constraints on origin and destination regions.

With the above description, the multi-objective optimization problem for the RFRRf technique is given as:

$$\min_{\delta\boldsymbol{\omega},\hat{\boldsymbol{u}}} \hat{f}(\delta\boldsymbol{\omega}), \hat{g}(\hat{\boldsymbol{u}}) \tag{5.7a}$$

$$\text{s.t.} \quad \sum_{i \in Z} \delta\omega_i^+ - \sum_{i \in Z} \delta\omega_i^- = 0 \tag{5.7b}$$

$$\sum_{j \in Z} \hat{u}_{ij} = \delta\omega_i^- \qquad \forall i \in Z \tag{5.7c}$$

$$\sum_{i \in Z} \hat{u}_{ij} = \delta\omega_j^+ \qquad \forall j \in Z \tag{5.7d}$$

$$\delta\omega_i^- \leq |V_i^+| \qquad \forall i \in Z \tag{5.7e}$$

$$\delta\omega_i^+ \leq \sum_{j \in Z} |V_j^+| \qquad \forall i \in Z \tag{5.7f}$$

$$\delta\omega_i^+ \cdot \delta\omega_i^- \geq 0 \qquad \forall i \in Z \tag{5.7g}$$

$$\delta\boldsymbol{\omega} := \delta\boldsymbol{\omega}^+ - \delta\boldsymbol{\omega}^-$$

$$\delta\boldsymbol{\omega}^+, \delta\boldsymbol{\omega}^- \in \mathbb{Z}_{\geq 0}^{|Z|}$$

$$\hat{\boldsymbol{u}} \in \mathbb{Z}_{\geq 0}^{|Z| \times |Z|}$$

$$\hat{g}(\hat{\boldsymbol{u}}) = \sum_{i,j} c_{ij} \hat{u}_{ij} \tag{5.7h}$$

$$\hat{f}(\delta\boldsymbol{\omega}) = F\left(\frac{3}{\pi} \sum_{i=1}^{n_z} \left(\omega_i^{sup} + \delta\omega_i\right) \frac{k(x, x_i^Z, h_i)}{h_i^2}, \quad \frac{3}{\pi} \sum_{i=1}^{n_z} \omega_i^{dem} \frac{k(x, x_i^Z, h_i)}{h_i^2}\right) \tag{5.7i}$$

where $\hat{\boldsymbol{u}}$ is a flow matrix with each element $\hat{u}_{ij}$ representing the number of idle vehicles repositioned from zone i to zone j and $\delta\boldsymbol{\omega}$ is a vector representing the overall change in the weight of a zone. The changes in the zone weights $\delta\boldsymbol{\omega}$ are broken into positive $\delta\boldsymbol{\omega}^+$ and negative $\delta\boldsymbol{\omega}^-$ contributions to the zone weights; thus, Eq. 5.7g makes sure that either of them are not simultaneously non-zero for each zone. $\boldsymbol{C} := (c_{ij}) \in \mathbb{R}_{\geq 0}^{|Z| \times |Z|}$ is the matrix of travelling costs between zone centroids. Eq. 5.7c and 5.7d guarantee that the total numbers of vehicles leaving a zone and entering other zones are in accordance with positive and negative changes to zone weights, respectively. Eq. 5.7e ensures that the negative changes to a zone weight (number of vehicles leaving the zone) are restricted by the number of idle vehicles available in the zone. On the contrary, the main purpose of the constraint on $\delta\boldsymbol{\omega}^+$ in Eq. 5.7f is to prune the search space as the total increase in a zone weight cannot be more than the total number of idle vehicles in the entire operating area.

For calculating the difference between supply and demand densities in Eq. 5.7i, the integral of squared deviation is used as it puts more importance on regions with high imbalance values than other metrics e.g. the integral of absolute deviation. Additionally, it significantly reduces the computational effort as described below:

$$\hat{f}(\boldsymbol{\delta\omega}) = a^2 \int \left[\sum_{i \in Z} \left(\underbrace{\omega_i^{sup} - \omega_i^{dem}}_{=\omega_i} + \delta\omega_i \right) \frac{k(x, x_i^Z, h)}{h_i^2} \right]^2 d\Omega \tag{5.8}$$

$$= \sum_{i \in Z} \sum_{j \in Z} \underbrace{(\omega_i + \delta\omega_i)(\omega_j + \delta\omega_j)}_{\text{constant for integration}} \int \frac{k(x, x_i^Z, h)}{h_i^2} \frac{k(x, x_j^Z, h)}{h_j^2} d\Omega \tag{5.9}$$

where $a = \frac{3}{\pi}$. With the definition of matrix $\boldsymbol{A} \in \mathbb{R}_{\geq 0}^{|Z| \times |Z|} : (A)_{ij} = \frac{1}{h_i^2 h_j^2} \int k(x, x_i^Z, h) k(x, x_j^Z, h) d\Omega$, the above equation can be written as:

$$\begin{aligned} \hat{f}(\boldsymbol{\delta\omega}) &= a^2 (\boldsymbol{\omega} + \boldsymbol{\delta\omega})^T \boldsymbol{A} (\boldsymbol{\omega} + \boldsymbol{\delta\omega}) \\ &= a^2 \left(\boldsymbol{\omega}^T \boldsymbol{A} \boldsymbol{\omega} + (2\boldsymbol{\omega}^T + \boldsymbol{\delta\omega}^T) \boldsymbol{A} \boldsymbol{\delta\omega} \right) \\ &= a^2 (2\boldsymbol{\omega}^T + \boldsymbol{\delta\omega}^T) \boldsymbol{A} \boldsymbol{\delta\omega} + Const \end{aligned} \tag{5.10}$$

The constant term $Const = a^2 \boldsymbol{\omega}^T \boldsymbol{A} \boldsymbol{\omega}$ can be ignored for the optimization problem. The major advantage of the formulation arises from the matrix $\boldsymbol{A}$ as it can be easily preprocessed for fixed zones (or fixed centers of RFs) using different numerical methods. The dissertation uses the midpoint rule for this purpose. Thus, $\boldsymbol{A}$ is given as:

$$\boldsymbol{A} = \begin{bmatrix} \frac{1}{h_1^4} \sum_{l,m} \hat{k}^2(x_1^Z, h_1) & \dots & \frac{1}{h_1^2 h_{n_z}^2} \sum_{l,m} \hat{k}(x_1^Z, h_1) \hat{k}(x_{n_z}^Z, h_{n_z}) \\ \dots & \dots & \dots \\ \frac{1}{h_1^2 h_{n_z}^2} \sum_{l,m} \hat{k}(x_{n_z}^Z, h) \hat{k}(x_1^Z, h) & \dots & \frac{1}{h_{n_z}^4} \sum_{l,m} \hat{k}^2(x_{n_z}^Z, h) \end{bmatrix} \Delta x \, \Delta y \tag{5.11}$$

where $\hat{k}(x_i^Z, h)$ is the evaluation of the RF $k(x, x_i^Z, h)$ on a discretized two dimensional grid with step sizes Δx and Δy. It also should be noted that $\boldsymbol{A}$ is a symmetric matrix. Thus, $\boldsymbol{A}$ can be calculated faster by only calculating the upper or lower triangular elements of the matrix and calculating the full matrix from it. Secondly, an element of $\boldsymbol{A}$ could only be a non-zero value when there is an overlap between the two RFs. For the RF used in the dissertation, i.e. triangular function with euclidean norm (Eq. 4.3), this could be quickly tested using the following condition:

$$h_i + h_j > d(x_i^Z, x_j^Z) \tag{5.12}$$

where $d(x_i^Z, x_j^Z)$ is the euclidean distance between the two RF centers. Thus, the condition in Eq. 5.12 could be tested beforehand to ensure that the concerned element of $\boldsymbol{A}$ is non-zero. Eq. 5.12 also shows that if the kernels have high bandwidths within a small area (an example is the Manhattan area with free-flow speed as shown in Figure 4.2a), there will be a lot of overlap between kernels and the sparsity of $\boldsymbol{A}$ will decrease. This will cause an increase in the number of quadratic terms in the objective function (Eq. 5.10), increasing the overall problem complexity.

It should also be noted that as the bandwidths of all regions approach zero (no overlap of the kernel functions), $\boldsymbol{A}$ becomes the unity matrix and Eq. 5.10 reduces to:

$$\hat{f}(\boldsymbol{\delta\omega}) = a^2 \sum_{i=1}^{n_z} (\omega_i + \delta\omega_i)^2 \tag{5.13}$$

In this form, the regions are assumed to be independent of each other. However, it is still quite different from other repositioning formulations that use the balancing of regions as constraints while minimizing the repositioning VKT [PAVONE et al., 2012; ALONSO-MORA, SAMARANAYAKE, et al., 2017]. Because of the multi-objective formulation in Eq. 5.7 and the square of zone weights in Eq. 5.13, the obtained repositioning problem will prefer balancing the regions with higher imbalances. This also allows the repositioning of vehicles to far-off regions, in contrast to the traditional limits of repositioning only to the closest regions as described in section 5.1.

5.4.3. Repositioning Formulations with Zone Restrictions

Since the vehicle supply and demand density terms are more general than surplus and deficiency of vehicles in individual regions, the flow of vehicles in the RFRRf formulation was kept general without any regional restrictions. This means that the idle vehicles could be repositioned to and from any region irrespective of local imbalances which allows the minimization of overall density as much as possible. However, since the dissertation uses integral squared deviation to simplify and reduce the computational effort (Eq. 5.10), even small differences in the density objective could be overemphasized during the optimization process. This can lead to excessive repositioning of vehicles. Secondly, the regional demands vary significantly during night and day hours. For certain times of the day, the majority of regions may not be imbalanced. However, the RFRRf formulation would still try to distribute the surplus of available vehicles throughout the city to decrease the overall ID and increase the areas reachable by vehicles. Both of these reasons can lead to increased repositioning of vehicles without a significant increase in overall performance. Thus, this section presents formulations that restrict the flow of vehicles from certain regions based on their local imbalance to lower down the above effects. These restrictions are put in two steps:

1. In the first modification of RFRRf, Eq. 5.7e is replaced by the following condition:

$$\delta\omega_i^- \leq \min(\max(0, \omega_i), |V_i^+|) \qquad\qquad \forall i \in Z \qquad (5.14)$$

   This constraint restricts the negative changes in a zone (the outflow of vehicles) to be zero if the original zone weight is negative (i.e. deficient zone). Since this formulation only allows the outflow of vehicles from surplus zones (zones with positive weights), the formulation is referred to as RFR with regions and positive zones flow (RFRRp).

2. The RFRRp restricts the outflow of vehicles from the deficiency to surplus zones, but there may still be excessive repositioning. The repositioning can happen from surplus to surplus zones to reduce the imbalance density. On the one hand, it may be beneficial to bring a higher number of vehicles to a zone than the demand expected if demand and supply forecasts contain uncertainties; on the other hand, this may cause unnecessary repositioning, especially when the majority of zones are not balanced. Therefore, the following approach restricts the overall formulation further by restricting the flow of vehicles to only from surplus to deficiency zones. This would also simplify the formulation as the number of variables and the required matrix sizes in the formulation would decrease as given below:

$$\min_{\boldsymbol{\delta\omega}, \boldsymbol{\hat{u}}} \hat{f}(\boldsymbol{\delta\omega}), \hat{g}(\boldsymbol{\hat{u}}) \tag{5.15a}$$

$$\text{s.t.} \quad \sum_{i \in Z} \delta\omega_i = 0 \tag{5.15b}$$

$$\sum_{j \in Z^-} \hat{u}_{ij} = -\delta\omega_i \qquad \forall i \in Z^+ \tag{5.15c}$$

$$\sum_{i \in Z^+} \hat{u}_{ij} = \delta\omega_j \qquad \forall j \in Z^- \tag{5.15d}$$

$$0 \geq \delta\omega_i \geq -\min(\omega_i, |V_i^+|) \qquad \forall i \in Z^+ \tag{5.15e}$$

$$0 \leq \delta\omega_i \leq -\omega_i \qquad \forall i \in Z^- \tag{5.15f}$$

$$\boldsymbol{\delta\omega} \in \mathbb{Z}^{|Z|}$$

$$\boldsymbol{\hat{u}} \in \mathbb{Z}_{\geq 0}^{|Z^+| \times |Z^-|}$$

Notice the decrease in the size of flow variables matrix $\boldsymbol{\hat{u}}$ and travelling costs matrix $\boldsymbol{\hat{C}}$ from $\mathbb{Z}^{|Z| \times |Z|}$ in RFRRf (and RFRRp) to $\mathbb{Z}^{|Z^+| \times |Z^-|}$. The requirement for breaking the $\boldsymbol{\delta\omega}$ into positive and negative parts are also removed since the changes in zone weights of deficiency and surplus zones could be only positive (Eq. 5.15f) and negative (Eq. 5.15e), respectively. Thus, conditions Eq. 5.15c and 5.15d directly use $\boldsymbol{\delta\omega}$ to ensure that the changes in zone weights are consistent with the number of idle vehicles. This further prunes the solution space according to current weights. Condition 5.15b guarantees the conservation of changes in zone weights.

Since this formulation is closer to the traditional understanding of repositioning — the vehicles only flow from surplus to deficiency regions — this approach will be referred to as RFRR.

In the rest of the dissertation, all the three variants of ID based repositioning (RFRR, RFRRp and RFRRf) are generally referred to as RFR methods. If any specific method out of the three variants is meant, the dissertation will use the name of the corresponding variant.

5.4.4. Implemented Optimization Approaches

The presented RFR methods (Eq. 5.7) are multi-objective optimization problems, which can be solved in multiple ways. This section first presents the solution approach where the highest priority is given to the density objective. Next, it presents a weighted sum approach that finds the Pareto front of the problem.

Lexicographic Method

In a multi-objective optimization problem, if the priority of the objective functions is known beforehand then the lexicographic method can be used for efficiently solving the problem [K. CHANG, 2015]. One major advantage of the method is that it gives Pareto optimal solutions without the need of scaling the individual objectives. In a lexicographic method, the problem is solved separately for each objective function. First, it is solved for the highest priority objective. Then, the already solved objective with the optimal value is put as a constraint while solving the problem for the next objective. Since separate variables $\boldsymbol{\delta\omega}$ and $\boldsymbol{\hat{u}}$ are used for the density and VKT objectives, respectively, first the optimization problem for $\hat{f}(\boldsymbol{\delta\omega})$ can be solved for optimal $\boldsymbol{\delta\omega}$ and then these optimal

values can be put as constraints on $\delta\boldsymbol{\omega}$ when solving for $\hat{g}(\hat{\boldsymbol{u}})$. The dissertation calls this approach prioritized balanced density (PBD).

Normalized Weighted Sum

The weighted sum method is a well-known solution method for multi-objective problems. In this method, the objective functions are multiplied by constant weights for the relative importance of individual objectives. The assigned weights help to move over Pareto front solutions. However, as the units and the magnitude of the individual objectives may vary for different problem instances, it requires a meaningful scaling of the individual objectives. The dissertation uses the well-known approach of scaling the objectives using nadir and utopian values [GRODZEVICH and ROMANKO, 2006]. Thus, the multi-objective function in Eq. 5.15a can rewritten as:

$$\min_{\delta\boldsymbol{\omega},\hat{\boldsymbol{u}}} \gamma \frac{\hat{f}(\delta\boldsymbol{\omega}) - \hat{f}_{utopia}}{\hat{f}_{nadir} - \hat{f}_{utopia}} + (1-\gamma)\frac{\hat{g}(\hat{\boldsymbol{u}}) - \hat{g}_{utopia}}{\hat{g}_{nadir} - \hat{g}_{utopia}} \tag{5.16}$$

where $\gamma \in [0,1)$ is a constant weight for the relative importance of density and distance objectives. The value of γ cannot be perfectly 1 as this makes the repositioning formulation only have the density as the objective, leading to multiple optimal solutions with random distance objective and flow variables. $\gamma = 0$, on the other hand, always has the unique solution of no repositioning. The MSP can choose between different values of γ according to the willingness to invest in the extra VKT for repositioning.

The utopian and nadir points are obtained from the Pareto optimal set of solutions. This set is usually obtained by optimizing the individual objectives with the original constraints. The lower bound from the set of optimal solutions forms the utopian point and the upper bound forms the nadir point. The lower bound is called utopian because generally such a point for an objective is not achievable while considering multiple objectives.

In the current problem formulation, it is sufficient to consider two extreme Pareto solutions for finding utopian and nadir points. The distance objective $\hat{g}(\hat{\boldsymbol{u}})$ is minimum when there is no movement of vehicles, i.e. no repositioning: $\hat{g}_{utopia} = 0$. This solution results in the worst density objective along the Pareto front and is equal to the initial value i.e. $\hat{f}_{nadir} = \hat{f}_{initial}$. Similarly, the other extreme Pareto optimal solution with minimum density objective is produced when the PBD is used. If $\hat{f}_{pbd}$ and $\hat{g}_{pbd}$ denote the objective values of density and distance objective after solving with PBD approach, then Eq. 5.16 is given as:

$$\min_{\delta\boldsymbol{\omega},\hat{\boldsymbol{u}}} \gamma \frac{\hat{f}(\delta\boldsymbol{\omega}) - \hat{f}_{pbd}}{\hat{f}_{initial} - \hat{f}_{pbd}} + (1-\gamma)\frac{\hat{g}(\hat{\boldsymbol{u}})}{\hat{g}_{pbd}} \tag{5.17}$$

Thus, for solving the RFR methods using the weighted sum approach, the first step is always to solve the problem using PBD approach. In the second step, Eq. 5.17 is used as the objective function in the RFR method and solved using MIP solver.

5.5. Experiments on Static Problem Instances

Before evaluating the suggested methods in a dynamic simulation, it is crucial to understand the basic principles and differences on static instances. This helps in analyzing the long-term dynamic simulation results. Therefore, this section presents the numerical experiments on static problem instances derived from NYC data.

5.5.1. Static Problem Instances

The static problem instances are derived from the same user data as described in section 3.2.2. The current section aggregates the data using cell sizes Δs_{cell} of 1000 m As discussed in section 5.1, the geographical distribution of customer origin and destination differ significantly between morning and evening hours. For a detailed analysis of the repositioning methods, it is necessary to study their outputs for different scenarios. Therefore, the static problem instances are generated by taking the NYC data trips from two time periods: 9 to 9:30 am and 6 to 6:30 pm on 6 June 2016. The positions of idle cars are generated by randomly taking the customer destinations from half an hour before 9 am and 6 pm, respectively. The RF bandwidths are calculated using the $AP_{(90,90)}$ as described in section 4.2.4.

5.5.2. Integral KPIs for Repositioning Decisions

For a detailed analysis of the repositioning methods on static problem instances, the section also uses some additional KPIs to study the potential impact of repositioning decisions. The KPIs uses the integral of the region-based ID function to represent the spread of positive and negative regions as single numbers. Since the $\bar{p}_Z$ (Eq. 5.6) estimates the expected reachability density using regions, the integrals of positive and negative regions of $\bar{p}_Z$ can quantify the area expected to be with surplus and deficiency of vehicles, respectively. These integrals are described as:

$$K_{\geq 0}(\phi) = \int_{\Omega^+} \bar{p}_Z(x; \phi) d\Omega \qquad \Omega^+ = \{x \in \Omega : \bar{p}_Z(x; \phi) \geq 0\} \qquad (5.18)$$

$$K_{\leq 0}(\phi) = \int_{\Omega^-} \bar{p}_Z(x; \phi) d\Omega \qquad \Omega^- = \{x \in \Omega : \bar{p}_Z(x; \phi) \leq 0\} \qquad (5.19)$$

$$\bar{p}_Z(x; \phi) = \frac{3}{\pi} \sum_{i=1}^{n_z} \frac{\phi_i}{h_i^2} k(x, x_i^Z, h_i) \qquad \text{(Spatial density of imbalance)} \qquad (5.20)$$

where $\phi_i \in \{\omega_i, \omega_i + \delta\omega_i\}$ can be the pre- or post-repositioning weight for zone $i \in Z$. These KPIs quantifies the spread and the intensity of the positive and negative values in the ID function. In principle, for a high AMoD performance, it is important that $K_{\leq 0}$ is close to zero since it would mean that most of the customers have access to a sufficient number of AMoD vehicles and could be assigned an AMoD vehicle. Contrarily, a highly negative value of $K_{\leq 0}$ would mean that either a big portion of the operation area has a deficiency of vehicles or there are smaller areas with high vehicle deficiency (a concentration of customers and not enough vehicles). Nevertheless, this would imply that these areas require more vehicles. Similarly, a high $K_{\geq 0}$ would mean that customers in large parts of the operation area can access multiple AMoD vehicles or a high concentration of vehicles in some areas. A repositioning method would relocate the vehicles from these excess vehicle areas to areas with a deficiency of vehicles, causing both $K_{\geq 0}$ and $K_{\leq 0}$ to decrease in magnitude. Note that it is possible — even desirable — that $K_{\leq 0}$ is close to zero while $K_{\geq 0}$ is a highly positive number. In this case, the repositioning method should not reposition vehicles since most customers already have access to AMoD vehicles, and a high $K_{\geq 0}$ means that the VCO has a large number of vehicles that it can assign to customers.

Due to the above reasons, the section uses the difference of the above integrals to compare the efficiency of the repositioning decisions, given as:

$$\Delta K_{abs} = |K_{\geq 0}(\omega_i + \delta\omega_i)| + |K_{\leq 0}(\omega_i + \delta\omega_i)| - \left(|K_{\geq 0}(\omega_i)| + |K_{\leq 0}(\omega_i)|\right) \qquad (5.21)$$

Eq. 5.21 estimates the overall improvements in the supply-demand imbalance by the repositioning decisions. It quantifies the expected decrements in both positive and negative areas of the ID. It does not indicate whether the increment is due to a decrement in the surplus or deficiency areas of AMoD services. Nevertheless, it provides an improvement KPI to compare repositioning decisions.

It should also be noted that the above KPIs assumes that a single repositioning vehicle increases (or decreases) the zone by one as described in section 5.2.2. However, in reality, the repositioning vehicle takes some time to reach its destination zone.

5.5.3. Results

As mentioned in section 5.2.2, the repositioning problem is solved periodically, and only idle vehicles can be repositioned. These idle vehicles are only a fraction of the overall fleet, and their numbers keep changing. The rest of the (non-idle) fleet contributes to the regional weights of supply-demand imbalance. The section first analyses the repositioning behavior with a fixed number of idle vehicles, i.e., 1000 vehicles, and later studies their behavior as the number of idle vehicles changes.

General behavior with a fixed number of idle vehicles

The dissertation first compares the Pareto fronts of the repositioning methods. They are generated using increasing values of γ in Eq. 5.17. As shown in Figure 5.4, the repositioning problem has conflicting objectives which are apparent from the distance versus density plots. With varying values of γ, the solutions build Pareto fronts. As described in section 5.4.4, the PBD solutions form the endpoints of the Pareto front for each method. It also shows that Pavone's and min-distance methods are not on the Pareto fronts for the imbalance density objectives. RFRRf provides the best compromise between the two objectives. At first glance, all of the RFR methods provide a better compromise of the vehicle supply-demand imbalance and the repositioning distance.

Figure 5.4 also shows that the decrease in the imbalance density objective is not linearly related to the distance objective; initially, there is a drastic decrease in the imbalance objective with the increase in repositioning distance, and later on, it flats out for higher repositioning distances. This indicates that with lower values of γ the RFR methods preferred balancing the areas with the highest deficiency of vehicles. The main reason for this is the usage of integral squared deviation in Eq. 5.9. As the value of γ increases, the rate of decrease in supply-demand imbalance reduces, leading to a relative flattening of the Pareto fronts. In this area of the Pareto front, the RFR methods can lead to excessive repositioning without much improvement in the AMoD supply-demand imbalance. This is especially relevant for the RFRRf method that does not restrict the repositioning of vehicles based on local supply-demand imbalance and leads to the highest repositioning VKT. However, this phenomenon also depends on the specific repositioning problem instance, as the flattening of the Pareto front differs for 9 am and 6 pm cases. It is also important to note that in terms of the imbalance density objective, the min-distance method reduces the supply-demand imbalance more than Pavone's approach. It also causes lower repositioning VKT than Pavone's method. The main reason for this is that, unlike Pavone's method, the min-distance caters to the demands of individual regions, which lowers the imbalance objective more than Pavone's method.

Figure 5.5 shows the increase in repositioning VKT and the number of vehicles repositioned against γ. With $\gamma = 0$ no vehicles are repositioned as it causes the objective function to only focus on the distance objective. With an increasing value of γ, all the RFR methods try to minimize the objective for supply-demand imbalance while giving relative importance to the total repositioning distance. As the γ approaches 1, the RFR methods converge to their PBD solutions. Generally, the computation time for RFRR method is observed to be better than the other RFR variants since the RFRR only
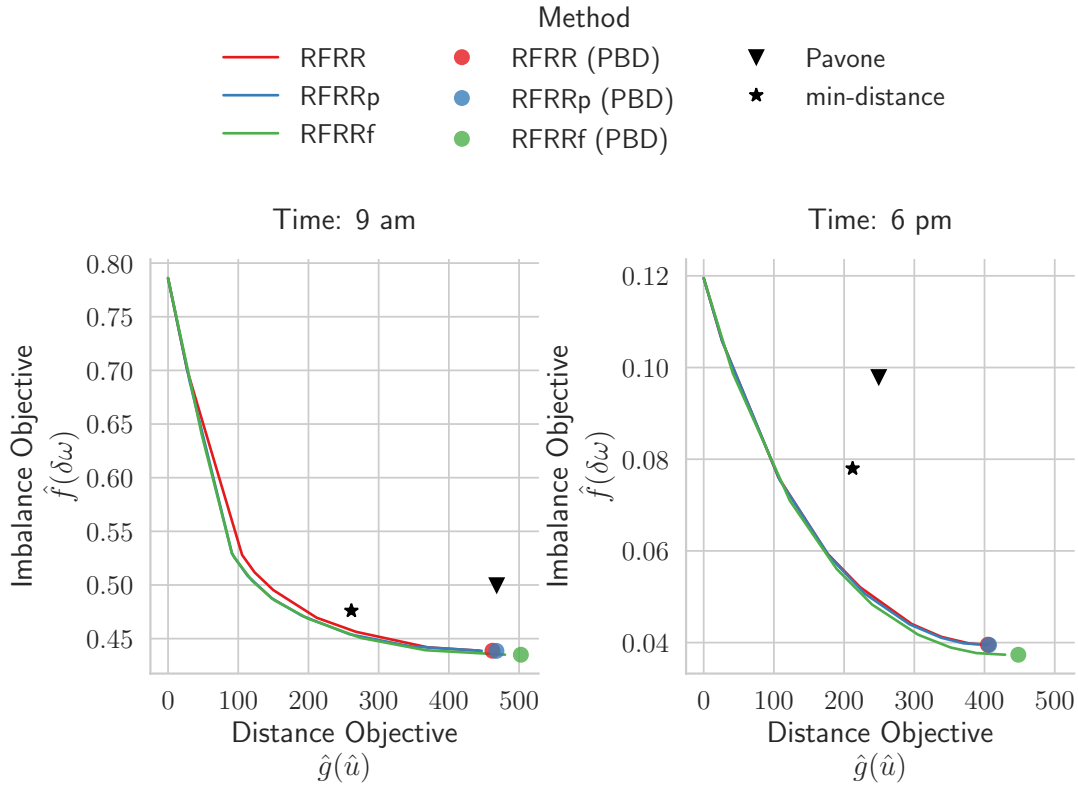
Figure 5.4.: Comparison of the Pareto fronts for the studied static cases with 1000 idle vehicles. The Pareto front is calculated using the increasing value of γ from 0 to 0.99 in Eq. 5.17.

.

sends vehicles from surplus to deficiency regions (smaller search space). As γ approaches one, the problem becomes difficult to solve due to a drastic increase in search space. However, even for such a high value of γ, there is not much difference in the solution qualities of the weighted sum and PBD solutions. It should also be noted that as described in section 5.4.4, the PBD has to be calculated first for the normalized weighted sum approach. Thus, the consistently low computational time of PBD is also beneficial for the weighted sum approach.

Figure 5.5 also shows that both RFRRp and RFRRf solutions result in a smaller imbalance objective than the RFRR method for all values of γ. Removing the constraint of only sending idle vehicles from surplus to deficiency regions enables the RFRRp and RFRRf to distribute the idle vehicles better. Compared to Pavone and min-distance, Figure 5.5 also shows that all RFR methods produce lower imbalance objective.

To investigate the methods further, Figure 5.6 and Figure 5.7 visualize static problems and their solutions on the map for 9 am and 6 pm cases, respectively. First, the regional supply-demand imbalances of the two cases are almost opposite due to different travel patterns of morning and evening hours (section 5.1). Second, both Pavone's and min-distance methods leave certain regions to be highly imbalanced; Pavone's method equally distributes the excess vehicles, which results in not repositioning enough vehicles to fulfill the demand of vehicle deficit regions, while the min-distance method prefers the neighboring regions of the idle vehicles to reduce the overall imbalance. This is apparent by comparing the pre-and post-repositioning weights of the regions for Pavone's and min-

(a) 9 am case

(b) 6 pm case
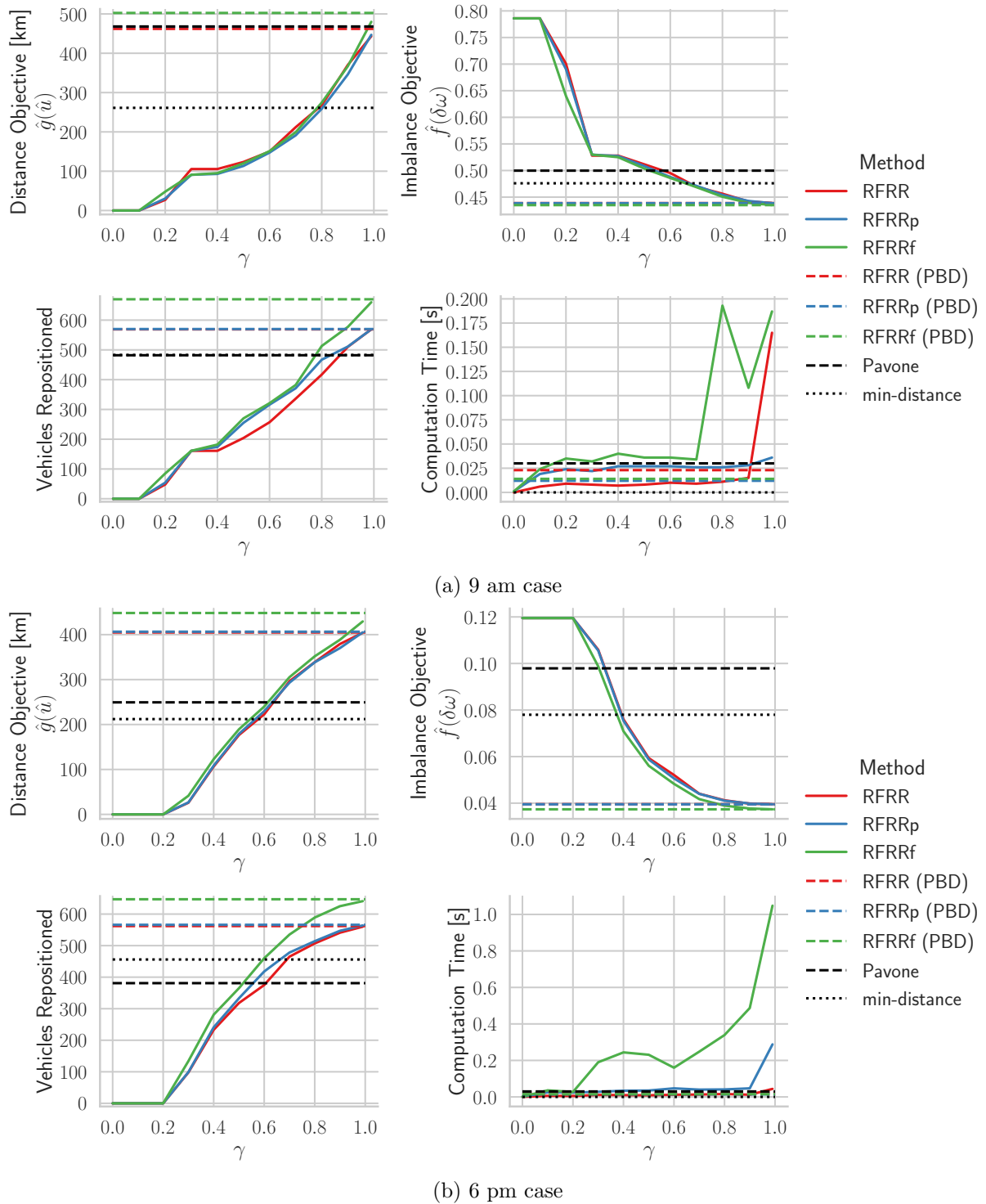
Figure 5.5.: Impact of the increasing values of γ. The computation time excludes the time required to set up the optimization problem. Additionally, for the RFR methods, it also excludes the time required for the PBD step, which is plotted separately in the figure.
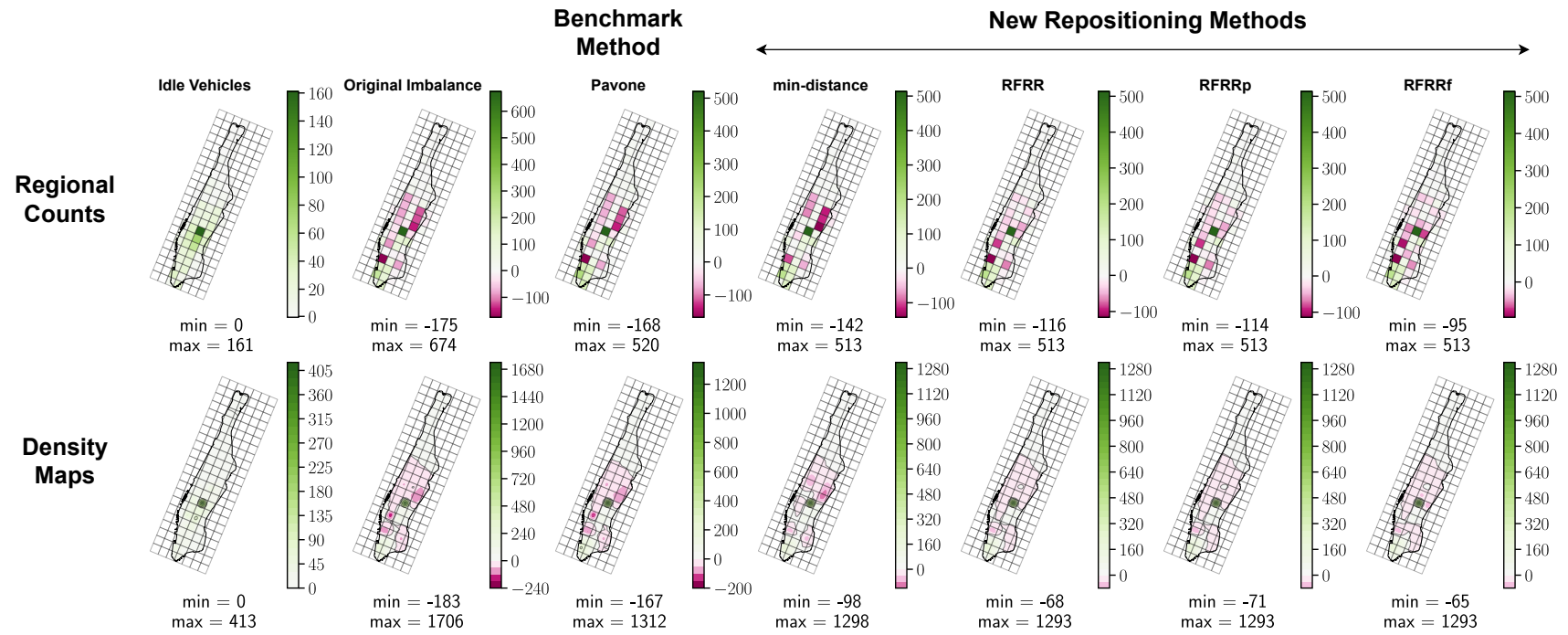
Figure 5.6.: Results of various repositioning methods for 9 am static case. The first row shows the zone counts while the second row shows the heat map calculated using zone-based ID. For ease of visualization, the values of ID are multiplied by 10^6. The RFR methods are solved using PBD approach.
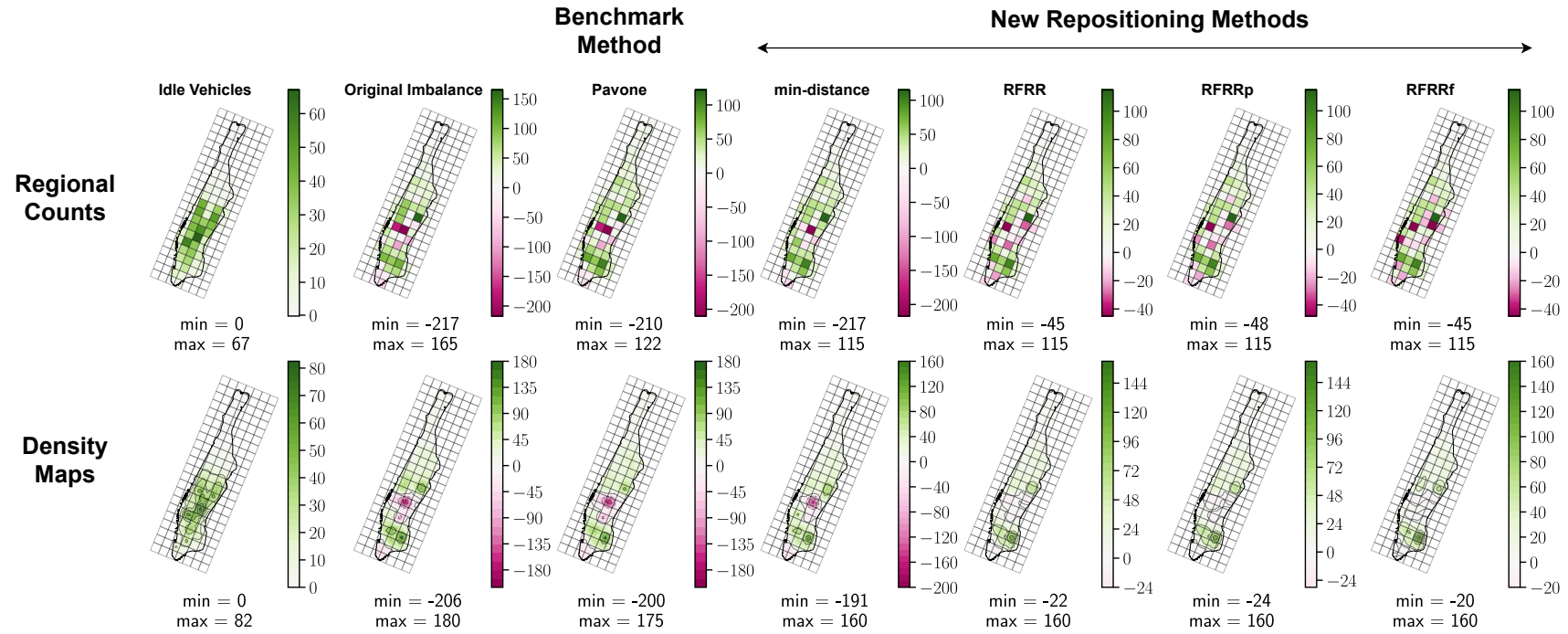
Figure 5.7.: Results of various repositioning methods for 6 pm static case. The first row shows the zone counts while the second row shows the heat map calculated using zone-based ID. For ease of visualization, the values of ID are multiplied by 10^6. The RFR methods are solved using PBD approach

distance methods for 9 am and 6 pm scenarios. This behavior also leaves bigger areas (considering the reachability) to remain imbalanced, as visible from the ID values of both Pavone's and min-distance methods.

Compared to the above, RFR methods focus on minimizing the imbalances in bigger areas, and thus, send more vehicles in a way that covers bigger regions. Mathematically, this is due to the integral of the squared deviation in Eq. 5.9 and the inclusion of ID in the repositioning formulation; the former helps to prefer areas with higher imbalances for repositioning and the latter allows to consider bigger areas via RFs. Since RFRRf does not use any regional constraints, it reduces the imbalance of the biggest areas possible, followed by RFRRp and RFRR methods. This is verified by comparing the magnitudes and the covered areas of the three methods for 9 am and 6 pm cases in Figure 5.6 and Figure 5.7, respectively. The figures also show the regional weights produced by the repositioning methods, which confirms the above assertion — instead of leaving multiple contiguous regions imbalanced, the RFR methods repositioned vehicles in a way that better covered various regions.

Impact of increasing idle vehicles

In the dynamic simulation setting, the number of idle vehicles in each call to repositioning keeps changing according to customer demands and VCO decisions. Therefore, Figure 5.8 and Figure 5.9 analyze the repositioning decisions with a growing number of idle vehicles. The new vehicles are generated using the same procedure as described in section 5.5.1. For consistency, the new vehicles are incrementally added to the problem instead of generating all vehicles from scratch. The most important observations are as follows:

- Looking at the imbalance objective in Figure 5.8, it is observed that for Pavone's, min-distance and RFRR methods, there is a critical number of vehicles at which the density objective is the least and the possible benefit from the applied repositioning method is maximum. This point will be referred to as vehicle saturation point (VSP). However, since RFRRp and RFRRf methods do not have strict regional constraints like RFRR, they continue to reposition a higher number of vehicles to improve the supply-demand balance among all areas, even if it marginally changes the density objective.

- In Figure 5.9, it is also observed that the RFR methods can change the absolute value of imbalance density ΔK_{abs} much more than Pavone's and min-distance methods, especially for smaller fleet sizes. This shows that the repositioned vehicles can cover bigger areas with the same fleet size. Interestingly, for min-distance and RFRR method, the fleet size with the largest absolute change ΔK_{abs} approximately coincides with the VSP, where these methods provide the highest benefits (measured via ΔK_{abs}) to the deficiency zones. Pavone's method causes the least decrease in ΔK_{abs} indicating the least improvement in AMoD service coverage among all the methods; rather, for the 6 pm case, the equal distribution of excess vehicles among all regions do not lead to a significant improvement in ΔK_{abs} even though a higher number of vehicles are repositioned as visible from Figure 5.8. For the RFRRp and RFRRf methods, the further decrease in ΔK_{abs} beyond VSP is due to additional balancing of the surplus areas (as discussed in the following points) which may lead to excess repositioning VKT and does not necessarily benefit the AMoD performance in a dynamic environment.

- For fleet sizes larger than the VSP, the imbalance objective in Figure 5.8 increases as the additional vehicles start to fulfill the regional demand and the need for additional repositioning decreases. This is also evident by comparing the values of total surplus $K_{\geq 0}$ and deficiency

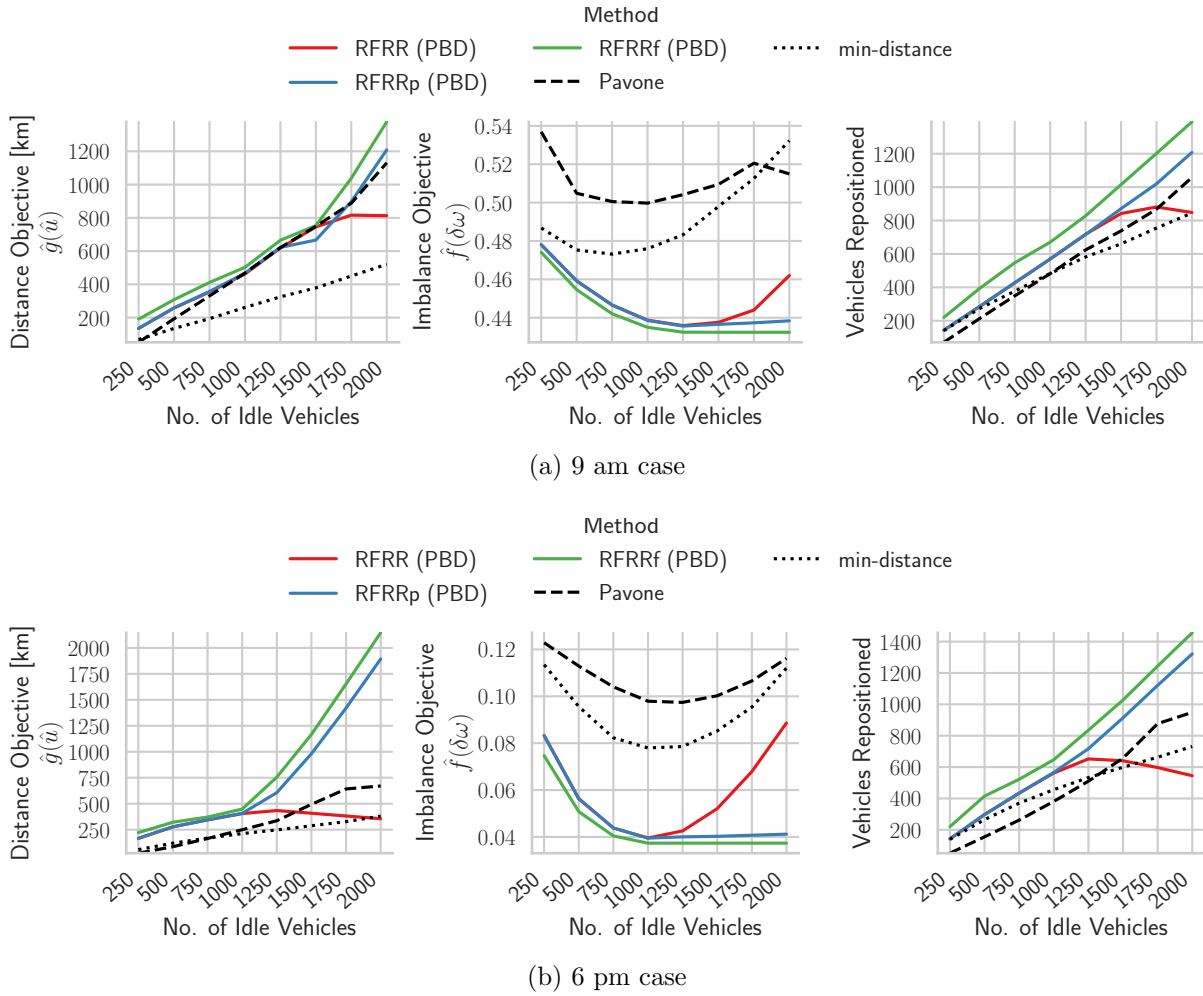(a) 9 am case

(b) 6 pm case

Figure 5.8.: Impact of the increasing number of idle vehicles.

$K_{\leq 0}$ for the original problem and the repositioning methods in Figure 5.9. Near VSP, $K_{\leq 0}$ gets close to zero, and the rate of decrease in deficiency $K_{\leq 0}$ is less than the rate of increase in surplus $K_{\geq 0}$.

- Compared to min-distance and Pavone's methods, the RFR methods perform much better for balancing the deficiency value $K_{\leq 0}$, but at the cost of comparatively higher VKT. The advantage of RFR methods is higher for fewer idle vehicles.

- Since integral square deviation is used for the density objective in Eq. 5.9, all RFR methods are equally sensitive to both surplus and deficiency, not just for deficiency. Therefore, it is observed that RFRRp and RFRRf keep repositioning more vehicles to balance the surplus density for the added vehicles even when $K_{\leq 0}$ has reached zero. The strategy of RFRRp and RFRRf might be beneficial when the accurate prediction of future requests is unavailable and the number of idle vehicles is small. However, if there are a lot of idle vehicles, this may lead to keeping many idle vehicles busy due to repositioning with overall non-profitable VKT.

- The RFRR method provides a good middle ground among all repositioning methods studied. Similar to mid-distance (and unlike RFRRp and RFRRf methods), the number of repositioned

(a) 9 am case

(b) 6 pm case
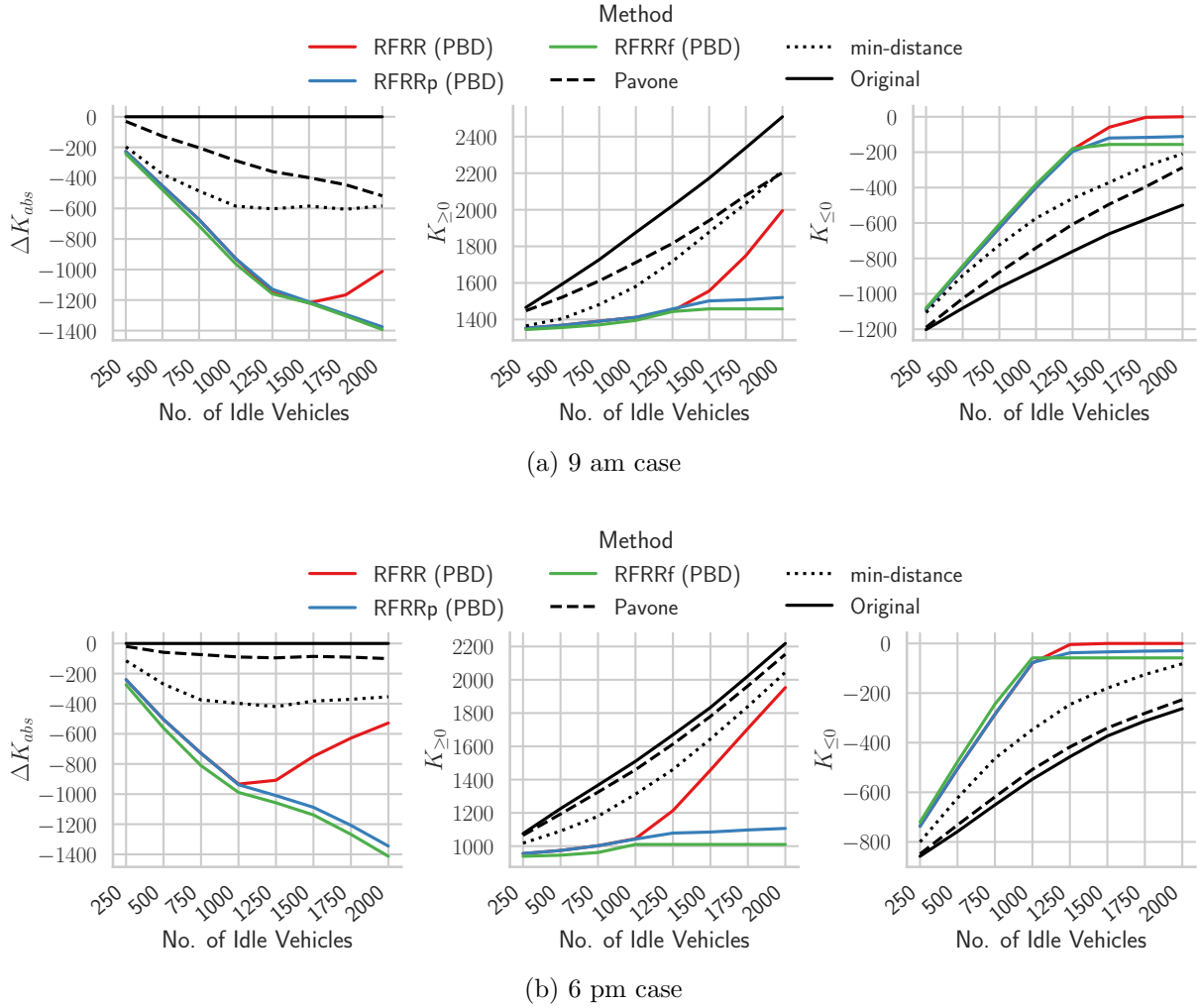
Figure 5.9.: Impact of the increasing number of idle vehicles on integral KPIs.

vehicles and the VKT decrease as the deficiency of regions ($K_{\leq 0}$) reaches zero. Secondly, for higher idle vehicles, the number of repositioning vehicles in RFRR is less than or equal to the min-distance and Pavone's methods, meaning fewer vehicles are busy with repositioning tasks. This would benefit the dynamic AMoD scenarios.

5.6. Experiments and Sensitivity Analysis in Agent-Based Simulation

The previous section showed the principle working of the introduced repositioning methods on static instances. However, the performance improvements of a repositioning algorithm can significantly differ from a static to a dynamic problem instance. The main reason is that in contrast to the assumption used while solving the static instances, the repositioning vehicles do not become immediately available in the destination zones; it takes some time for the repositioning vehicles to reach their destinations. Therefore, this section performs the sensitivity analysis of the most important parameters in the dynamic environment. Since the simulation setup and the introduced methods consist of many parameters, the section does the sensitivity analysis of the most important parame-

Parameter / Method	Symbol	Section for Sensitivity Analysis	Default Value / Strategy
Base fare	ζ	5.6.6	\$2.5 per customer
Distance-based variable fare	f^D	—	\$0.5 per km
Distance-based cost	c^D	—	\$0.25 per km
Fixed maintenance cost of vehicle	c^F	—	\$25 per vehicle per day
Maximum allowed waiting time of customers	ΔT_{max}	5.6.3	6 minutes
The time period used for scaling the travel times	ΔT_{scale}	—	30 minutes
Travel time scaling method	—	—	ASM (section 3.3)
Fleet size	—	5.6.6	3000 AVs
Vehicle assignment method	VCO	5.6.1, 5.6.3, 5.6.6	$\text{VCO}_{enroute}$ (section 3.1.3)
Batching period	ΔT_{batch}	—	30 seconds
RF bandwidth calculation method	—	5.6.5	$AP_{90,90}$ (section 4.2.4)
The repositioning period	ΔT_r	5.6.4	30 minutes
The forecast horizon for repositioning	ΔT_h	5.6.2	+30 minutes (Perfect forecast)
Grid cell size of regions in the operation area	Δs_{cell}	5.6.2	1000 m
The method used for vehicle supply estimate	ω_z^{sup}	5.6.2, 5.6.4, 5.6.6	$\omega_z^{sup(c)}$

Table 5.1.: The simulation configuration used in chapter 5.

ters. Table 5.1 summarizes the default simulation configuration used in the chapter along with the detail of the specific section where a certain parameter is varied. The base parameters remain the same as described in section 4.3 for consistency.

5.6.1. Relative Importance of Imbalance Objective and Repositioning Distance

Figure 5.5 in section 5.5.3 showed that an increase in the values of γ causes an increase in the repositioning vehicles and a decrease in the density objective. Ideally, this should lead to an improved AMoD performance for higher values of γ. However, the outcome can vary significantly in an agent-based simulation due to its dynamic nature. As mentioned in section 5.2.2, the repositioning problem assumes that a single repositioning vehicle changes the imbalance weight by one. In reality, a repositioned vehicle is not immediately available in the destination zone. Thus, a repositioning method's decrease in the supply-demand imbalance on static cases does not guarantee performance improvement in a dynamic setting. Therefore, the section first compares the impact of γ on the overall performance in the agent-based simulation.

As shown in Figure 5.10 and Figure 5.11, as γ increases, the performance of RFR methods tend to converge to the PBD solutions in a similar way as the static examples in Figure 5.5. For the RFRR method, the increase in γ also leads to a higher percentage of served customers $S_\%$ and monetary profit and a lower mean waiting time W_{mean}. The PBD solutions provide the best results for RFRR method. However, the absence of regional constraints in RFRRp and RFRRf methods
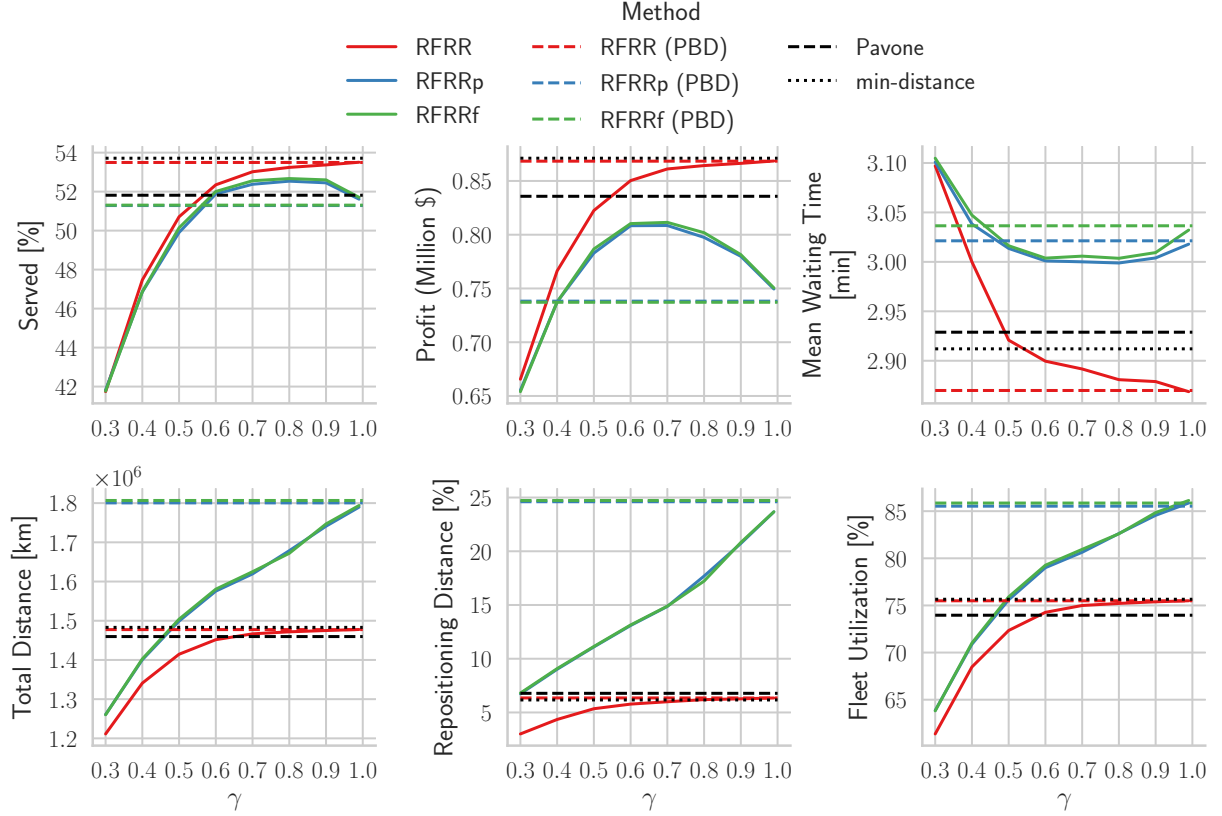
Figure 5.10.: The effect of increasing γ in the simulation environment with VCO$_{idle}$.

cause a significant increase in repositioning VKT (and cost), indicated by the total covered distance D_{total} and percentage of repositioning distance $D_\%^{repo}$. This keeps the idle vehicles busier with the repositioning tasks and can only serve a smaller number of customers than RFRR method. This is also evident from the high mean utilization rate $U_\%$ of RFRRp and RFRRf. Thus, even though RFRRp and RFRRf have higher $S_\%$ than Pavone's method, they produce a smaller monetary profit. Additionally, unlike RFRR method, RFRRp and RFRRf methods have their highest performance near γ of 0.7, after which the repositioning costs take over the benefits, and the monetary profit starts to decrease. The introduced min-distance method also produces a high AMoD performance, almost similar to the RFRR method. This shows that a major hindrance in tapping the full performance using Pavone's methods is the equal division of excess vehicles among regions which does not reposition enough vehicles to demand-intensive regions. In contrast, the min-distance minimizes the sum of the overall imbalance, which leads to a significant performance improvement over Pavone's method. The above performance improvement pattern is identical for both VCO types.

To further investigate the AMoD performances, Figure 5.12 shows the temporal graphs of the repositioning methods. The first observation is that repositioning idle vehicles resolves the performance degradation over time observed without repositioning in section 4.3.1. Another important observation is that the excessive repositioning of idle vehicles in RFRRf method mainly happens during the hours of low customer demand (i.e., during the night and the early morning). This is visible from the percentage of time AMoD fleet was busy performing repositioning tasks i.e., the $U_\%$ for repositioning. A major reason for this is that the PBD solution of RFRRf tries to distribute idle vehicles as much as possible to minimize the ID. During the low demand hours, there are a
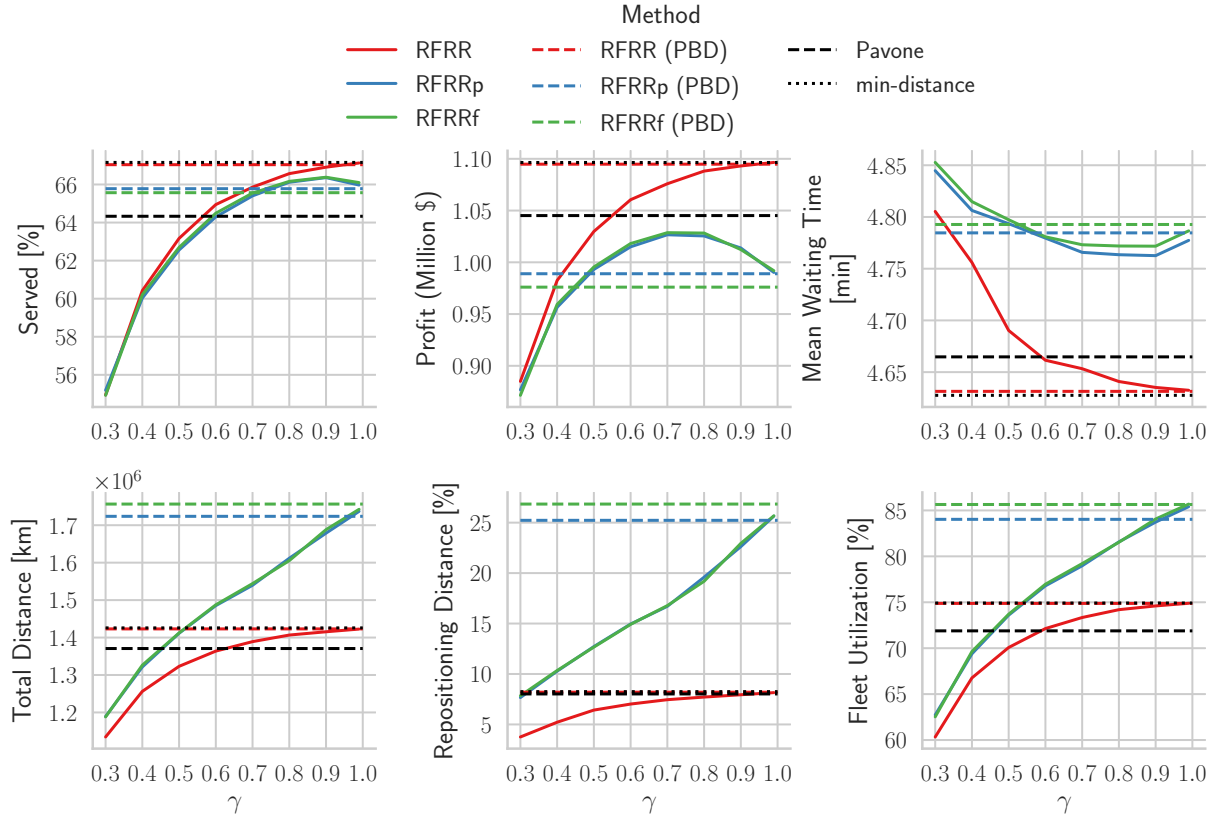
Figure 5.11.: The effect of increasing γ in the simulation environment with $\text{VCO}_{enroute}$.

lot of idle vehicles, which makes the RFRRf method reposition an excessive number of idle vehicles even if it fractionally reduces the ID. The same behavior was observed on static problem instances in Figure 5.8 for RFRRp and RFRRf methods: with a high number of idle vehicles, these two methods kept repositioning more vehicles for a marginal reduction in the density objective. Consequently, by using an appropriate value of γ, this excessive repositioning is reduced, increasing the RFRRf profit. During the daytime, the RFR and min-distance methods perform pretty similar to each other; however, RFRR and RFRRf methods sometimes have higher $S_\%$ than min-distance. Pavone's method shows the least $S_\%$, especially during the daytime.

While Figure 5.12 showed the temporal performances, Figure 5.13 and Figure 5.14 do further analysis using the ID based spatial metric ($\rho_{imb}(x)$) for the served and unserved requests, respectively. Unlike RFR methods where the future ID is estimated using aggregated values over regions, here the actual coordinates of customer and vehicle locations (the x in $\rho_{imb}(x)$) is used to calculate the ID of each VCO batch following the same procedure as in section 4.3.2. The first general observation is that the repositioning methods significantly moved the $\rho_{imb}(x)$ of the customer pickup points towards the positive side of $\rho_{imb}(x)$ compared to the scenario without repositioning. This means that due to the repositioning of idle vehicles, the customer pickup locations were close to availability points of AMoD vehicles, increasing the $S_\%$ for all time periods. The cause of the steady performance degradation without repositioning can also be observed in Figure 5.14; the values of $\rho_{imb}(x)$ steadily shift towards negative as the simulation progresses. This means that increasing customer pickup locations are not within the reachable distances of AMoD vehicles without repositioning. In contrast, the repositioning methods distribute idle vehicles throughout the city, increasing customers' access to the AMoD fleet
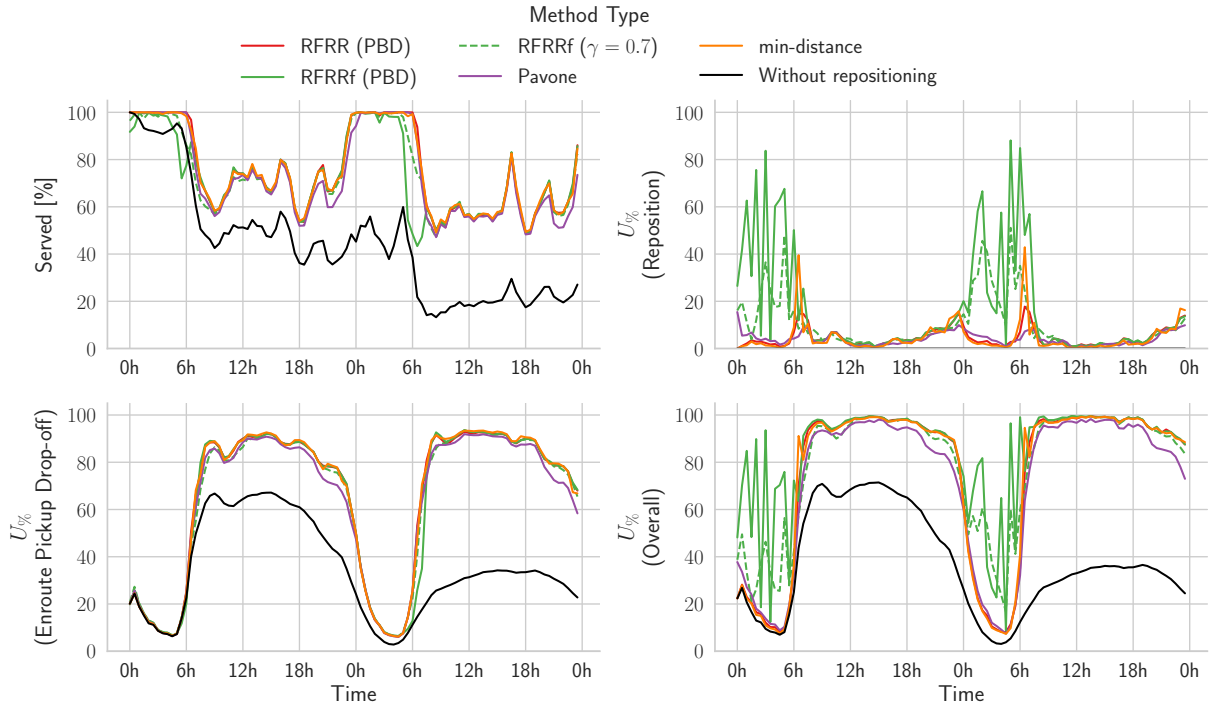
Figure 5.12.: The temporal performance of repositioning methods with 3000 vehicles and $\text{VCO}_{enroute}$

and helping to maintain high service quality.

It is also worthwhile to note that Figure 5.13 and Figure 5.14 used $\text{VCO}_{enroute}$ which, as described as in section 4.3.2 has a somewhat softer increase in $S_{\%}$ as the $\rho_{imb}(x)$ becomes positive. A similar outcome is seen in Figure 5.14 for the repositioning methods: some unserved customers also have positive $\rho_{imb}(x)$ though in lower value ranges than the served customers in Figure 5.13. Nonetheless, the repositioning methods have shifted the values of $\rho_{imb}(x)$ for almost all of the customers to the positive side, which would mean that the customers are within the range ΔT_{max} from the vehicle availability points. This increases the chances of vehicle assignments and leads to a higher $S_{\%}$ as evident from the temporal AMoD performances for the corresponding periods in Figure 5.12.

Figure 5.13 and Figure 5.14 also help to distinguish the performance of each repositioning method. The customers in Pavone's method have comparatively lower $\rho_{imb}(x)$ than other repositioning methods. The customers in the RFRR method have the highest values of $\rho_{imb}(x)$, which also lead to the highest $S_{\%}$ for all periods.

The effect of excessive repositioning in the RFRRf method during the night hours can also be observed in Figure 5.13; many vehicles remain busy with repositioning tasks causing the $\rho_{imb}(x)$ to be closer to zero than other repositioning methods, and thus, have the least $S_{\%}$. During the daytime, the performance of RFRRf and min-distance keeps changing with one method having higher $\rho_{imb}(x)$ and $S_{\%}$ than the other during different times of the day. This also shows that even though the min-distance method does not consider ID for repositioning, the goal of just sending enough vehicles to fulfill local demands with minimum VKT still leads to a high $\rho_{imb}(x)$ for the customers.
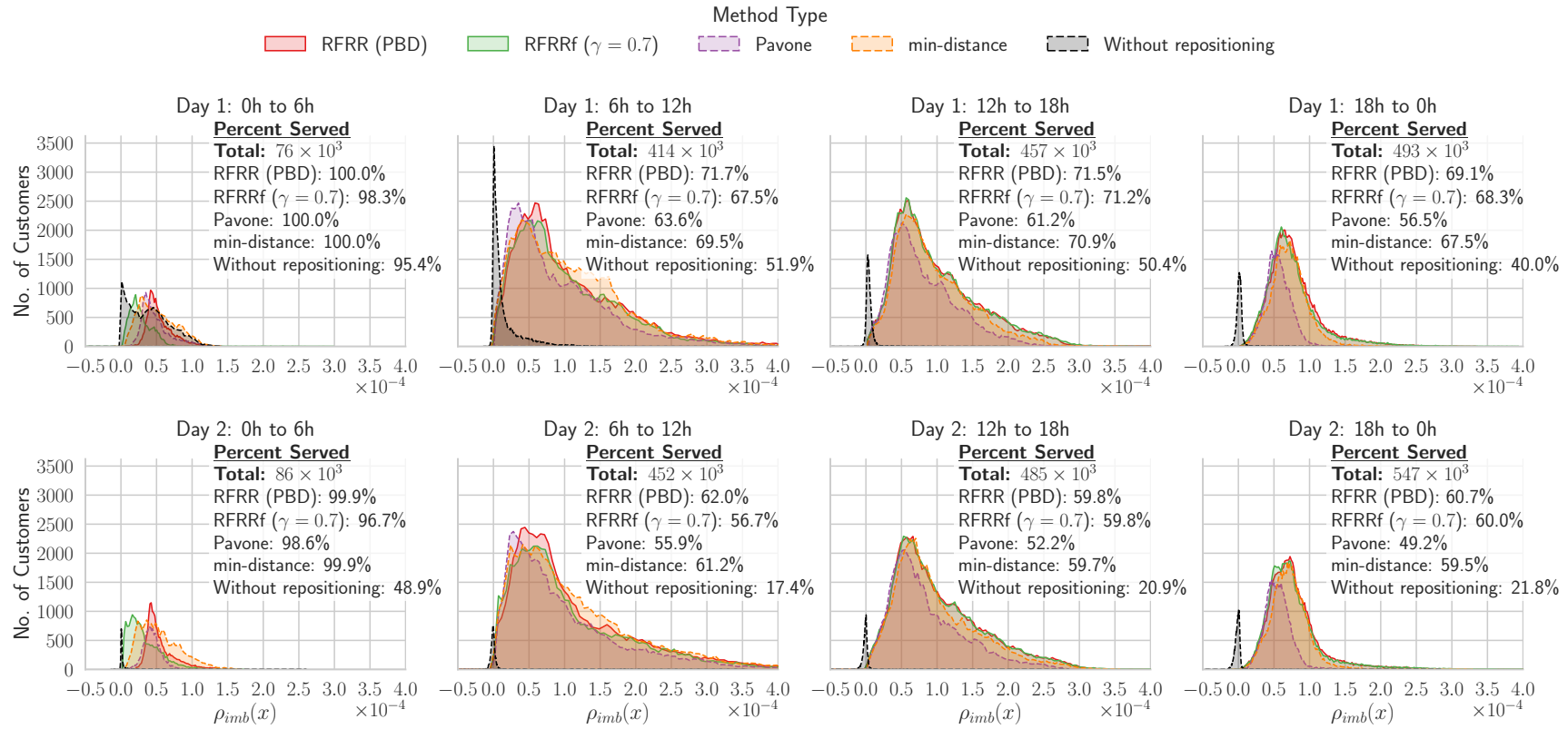
Figure 5.13.: The distribution of imbalance density values $\rho_{imb}(x)$ for the served customers. The simulations used $\text{VCO}_{enroute}$ and vehicle supply estimates using $\omega_z^{sup(c)}$ for repositioning.
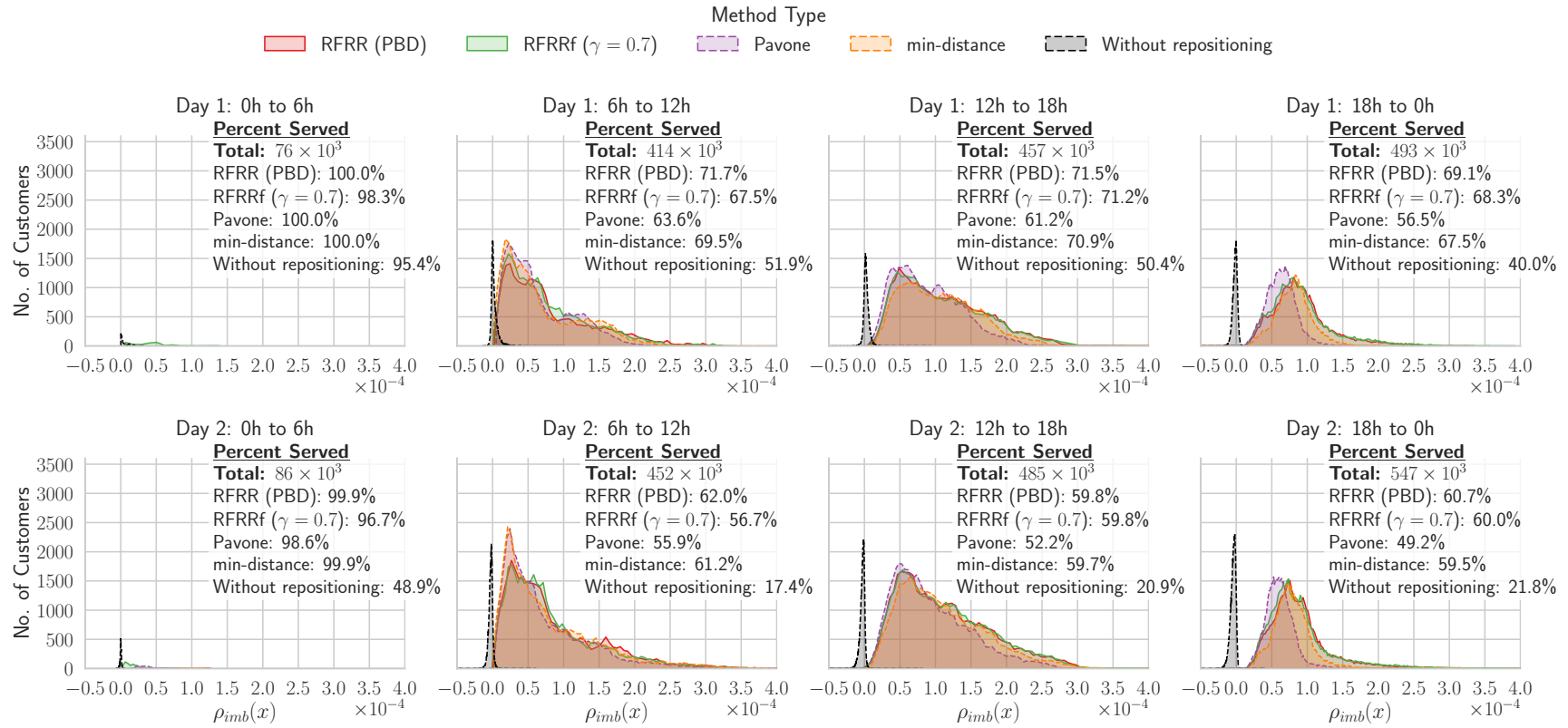
Figure 5.14.: The distribution of imbalance density values $\rho_{imb}(x)$ for the unserved customers. The simulations used $VCO_{enroute}$ and vehicle supply estimates using $\omega_z^{sup(c)}$ for repositioning.

5.6.2. Forecast Types and Accuracy

The previous section showed that even in the dynamic environment, the weighted sum solutions converge to PBD solutions as the value of γ reaches one. This section studies the impact of forecast accuracy and types on the performance of the repositioning methods.

The accuracy of the forecast is of utmost importance for any repositioning algorithm. The performance of a repositioning method under different forecast accuracy shows the robustness of the method. As described in section 5.2.2, the forecast accuracy depends on two types of information: spatial and temporal. A high spatial accuracy would estimate the customer pickup location on smaller regions, which in the dissertation corresponds to smaller Δs_{cell}). Similarly, the temporal accuracy of the forecast depends on the value of ΔT_h used. Furthermore, in the dissertation, a positive (perfect forecast) and smaller value of ΔT_h would mean that the forecast is temporally accurate and the customer requests will certainly be received within a short time interval. In contrast, a negative (imperfect forecast) and smaller value of ΔT_h would mean that it is not certain that a customer request will occur in the expected region, but if it occurs, it must be received within a short time interval. A larger ΔT_h adds a further challenge to the repositioning problem for both perfect and imperfect forecasts. On the one hand, it is beneficial to reposition more vehicles to increase fleet coverage, and on the other hand, it is not sure when exactly the vehicles will be needed due to large ΔT_h; sending more vehicles may keep the AMoD fleet busy with the repositioning tasks and might even decrease the performance.

Because of the above, the section focuses on evaluating the impacts of spatial and temporal accuracy of forecasts on the performance of repositioning methods. It does so in the following three steps. It first studies the effects of spatial accuracy Δs_{cell}, then the perfect and imperfect forecast's impact, and finally, the temporal resolution Δs_{cell}. For simplicity, the section only uses the PBD solutions for this purpose due to lower computational times. Additionally, since the overall AMoD service is an SDVRP, the forecast of the future state of vehicles also plays a significant role in the performance of repositioning methods (section 5.2.1). Therefore, in addition to the spatial and temporal accuracy, the section also focuses on the performance differences when various vehicle supply forecasts (i.e., $\omega_z^{sup(c)}$ and $\omega_z^{sup(v)}$) are used.

Spatial Accuracy Δs_{cell} and Perfect Forecast

The dissertation divides the operational area into regions using grids of cell size Δs_{cell}. A lower Δs_{cell} represents a higher spatial accuracy for the forecast. This section studies the impact of varying Δs_{cell} while keeping the ΔT_h fixed at +30 minutes. It only uses $VCO_{enroute}$ for this purpose as the objective is to understand the general behavior of the introduced methods when the spatial accuracy is varied.

Figure 5.15 shows the impact of Δs_{cell} with vehicle supply estimates using $\omega_z^{sup(c)}$ and $\omega_z^{sup(v)}$. First, all methods are significantly affected by varying values of Δs_{cell}. In general, the vehicle supply estimate using $\omega_z^{sup(c)}$ produces a higher monetary profit and serves more customers than $\omega_z^{sup(v)}$. This shows that considering the long-term supply of vehicles in the regional weights improves the performance of repositioning methods even though $\omega_z^{sup(c)}$ only estimated it via customer destinations in the data.

As for the impact of spatial accuracy, Pavone's method tends to perform the worst when the regions are small; a smaller Δs_{cell} results in a high number of regions and the equal distribution of excess vehicles in Pavone's method causes the high deficit regions not to receive sufficient repositioning vehicles. As Δs_{cell} increases, the total number of regions decreases, causing Pavone's method to send more vehicles to individual regions. This improves its performance as the deficit regions can now receive more vehicles. The best performance of Pavone's method is achieved at Δs_{cell} of 1000 m.
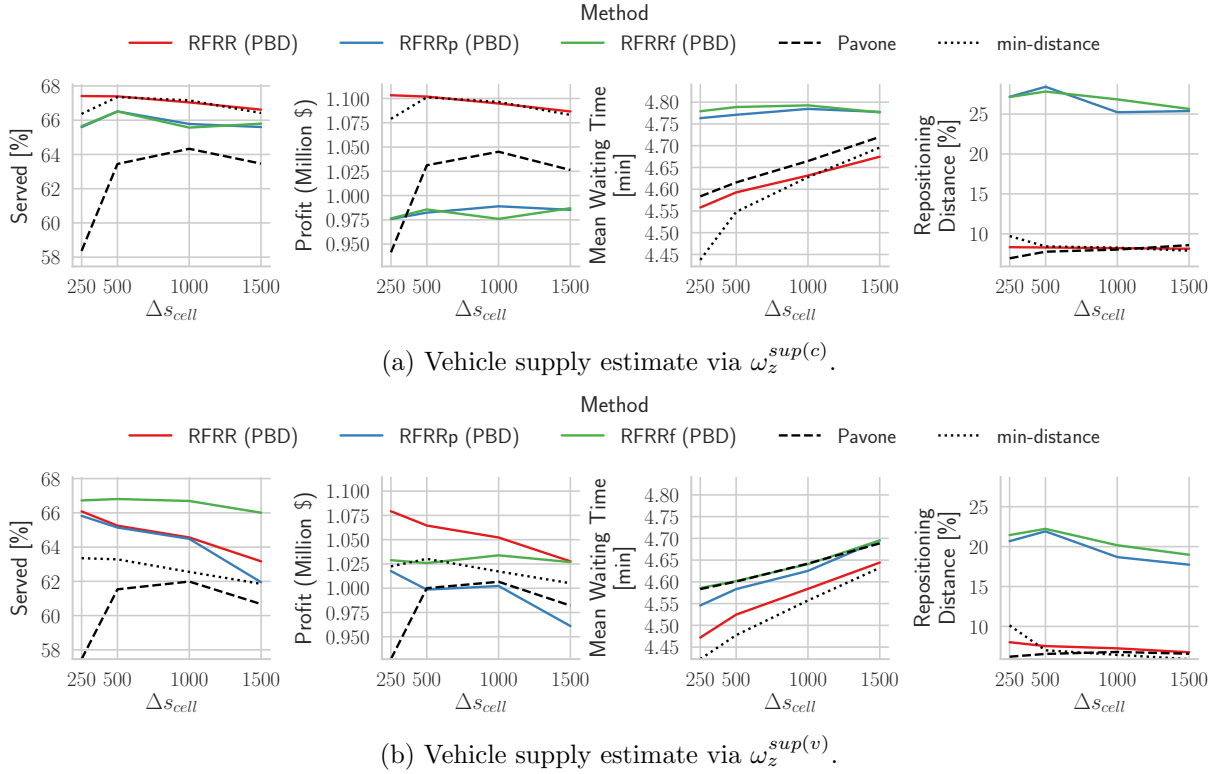
(a) Vehicle supply estimate via $\omega_z^{sup(c)}$.

(b) Vehicle supply estimate via $\omega_z^{sup(v)}$.

Figure 5.15.: The impact of spatial accuracy and supply estimations on the repositioning methods, ΔT_h of $+30$ minutes (perfect forecast) and $VCO_{enroute}$. A higher Δs_{cell} represents bigger regions and lower spatial accuracy.

As the regions grow even larger than 1000 m, the performance of Pavone's method decreases. A major reason for this is that the regions become so big that all customers within a region could not be reached within ΔT_{max} of 6 minutes from regional centers — the location where the vehicles are sent during repositioning. This is also indicated by the increasing mean waiting time (W_{mean}) with increasing Δs_{cell}. This degrades the overall performance of Pavone's method for larger regions.

As shown by Figure 5.15a, among all methods, RFRR provides the best performance with Δs_{cell} of 250 m and vehicle supply estimate using $\omega_z^{sup(c)}$, followed by min-distance method. Interestingly, the RFRR method performs significantly better than the min-distance method for Δs_{cell} of 250 m. The main reason is that small Δs_{cell} significantly increases the number of regions. The min-distance method prefers to balance the neighboring regions while trying to reposition as many excess vehicles as possible. For Δs_{cell} of 250 m, this leads to a high repositioning of idle vehicles due to a higher number of regions, as evident by $D_\%^{repo}$ in Figure 5.15. On the contrary, the usage of RDs makes the RFRR method aware of the regional sizes and proximity, causing it not to get affected by many regions and better distribute the repositioning vehicles. In fact, RFRR provides the best performance with Δs_{cell} of 250 m as a smaller grid offers a better estimate of the potential ID function (Eq. 5.6). For regions larger than 250 m, the min-distance and RFRR method provide almost similar performance. Similar section 5.6.1, the RFRRp and RFRRf lead to excess repositioning VKT with $\omega_z^{sup(c)}$ as shown in Figure 5.15a, causing a decrease in the overall performance.

Figure 5.15b shows that the performance of RFRR method is less susceptible to supply-side estimation than min-distance and Pavone's method, with RFRR providing the best performance for
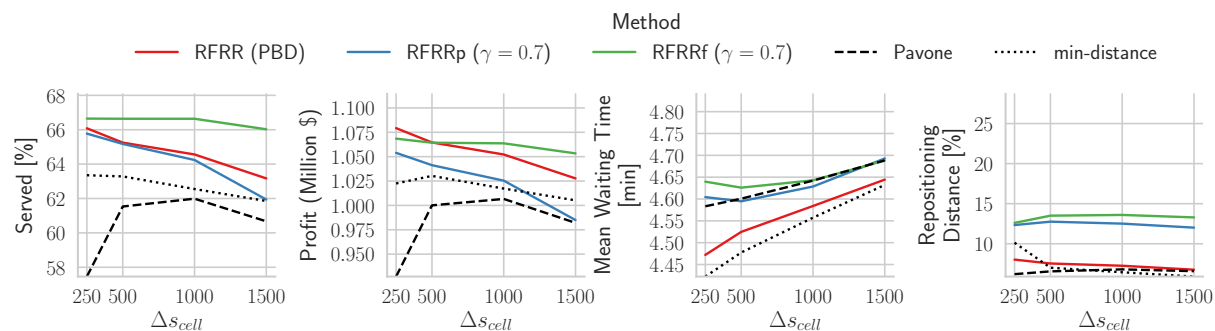
Figure 5.16.: The comparison of RFRRp and RFRRf methods with γ of 0.7 and vehicle supply estimate using $\omega_z^{sup(v)}$.

$\omega_z^{sup(v)}$. The main reason for the low susceptibility of the min-distance method is the assumption of independent regions (section 4.1); since each region is considered to be independent, the method heavily depends on the provision of a long-term vehicle supply estimate for better repositioning of idle vehicles. Otherwise, the method may wrongly estimate that some of the close-by regions have vehicle deficits and require repositioning, while in reality, some of these estimated vehicle deficits would be fulfilled by the vehicles dropping off customers in the region. Pavone's method would behave similarly by assuming fewer excess vehicles due to the absence of long-term vehicle supply estimates. Thus, it will reposition a smaller number of idle vehicles. In contrast to this, the RFRR has a significantly low drop in performance when $\omega_z^{sup(v)}$ is used for the vehicle supply estimate. Without a long-term vehicle supply estimate, the RFRR method would reposition the vehicles in a way that covers the most customer pickup locations due to the usage of RD. Thus, the RFRR method performs significantly better than Pavone's and min-distance methods with $\omega_z^{sup(v)}$.

Interestingly, with $\omega_z^{sup(v)}$, the performances of RFRRp and RFRRf are significantly improved: RFRRf serves the highest proportion of customers ($S_\%$) among all methods, followed by RFRR and RFRRp. This shows that without the long-term vehicle supply estimation, the RFRRf method distributed the AMoD vehicles in a way that increased the accessibility of more customers. However, the excessive repositioning in RFRRf method caused the overall profit of the RFRRf method to be still lower than RFRR method. A major reason for this performance gain of RFRRf with $\omega_z^{sup(v)}$ is the accuracy of vehicle supply estimate; since RFRRf tries to fully minimize the region-based ID without any restriction on surplus or deficit regions, the accuracy of the vehicle supply estimation is of crucial importance. The RFRRf method already tends to cause a high $D_\%^{repo}$ in an attempt to minimize the expected ID fully. Thus, an error-prone supply estimate in $\omega_z^{sup(c)}$ will cause it to reposition more vehicles (even if the reduction in ID is only marginal) with the wrong assumption that certain regions are already covered by AMoD vehicles. In contrast, even though $\omega_z^{sup(v)}$ does not have any term for the long-term forecast of vehicle supply, the short-term supply information is still highly accurate. The RFRRf tries its best to distribute AMoD fleet in a way that minimizes ID and, in turn, provides higher AMoD accessibility to a larger number of customer pickup points.

It should also be noted that this section only considered PBD based solutions. The excess repositioning could be reduced using the weighted sum approach with a lower value of γ for RFRRp and RFRRf methods. This way, the RFRRf method can further improve performance for $\omega_z^{sup(v)}$. Figure 5.16 shows such an improvement using γ of 0.7. The RFRRf significantly improves over all other methods for larger Δs_{cell}. It is also noteworthy that the performance of RFRRf method is least affected by the increasing value of Δs_{cell}. This also shows the benefits of minimizing the
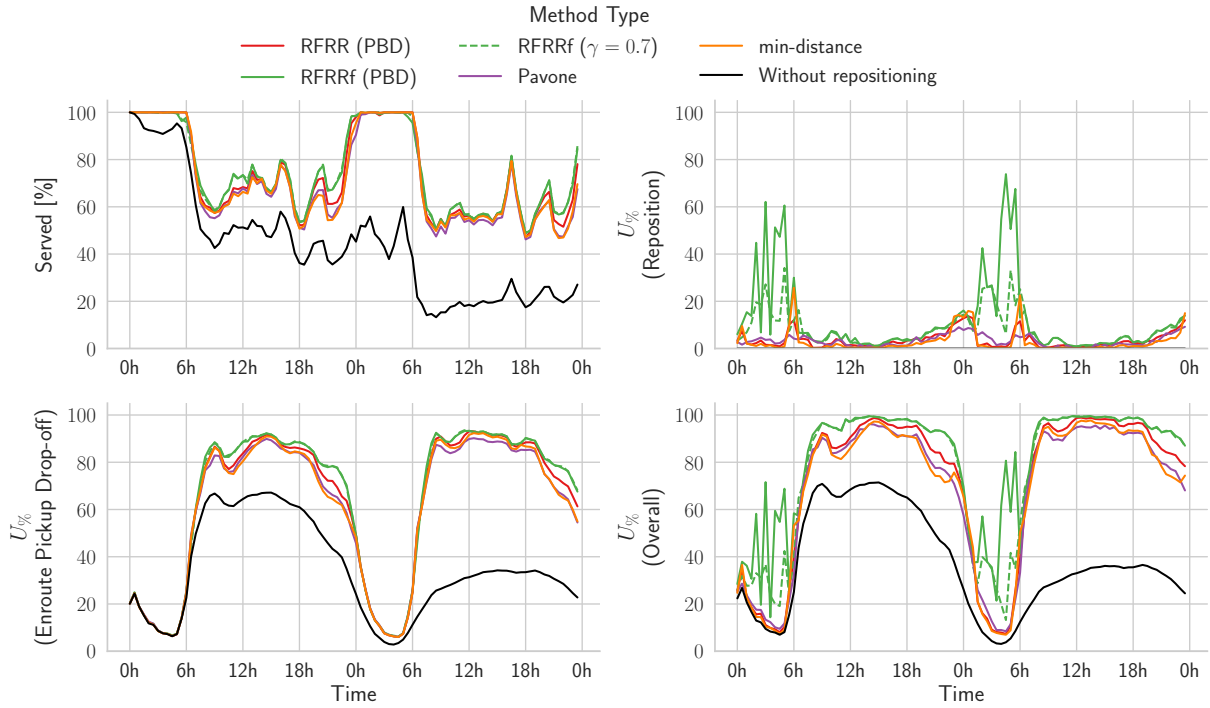
Figure 5.17.: The temporal performance with $\text{VCO}_{enroute}$, Δs_{cell} of 1000 m and vehicle supply estimates using $\omega_z^{sup(v)}$.

region-based ID for the purpose of repositioning. Figure 5.17 shows the temporal performance with $\omega_z^{sup(v)}$. The general behavior of the RFRRf method is the same as described in section 5.6.1; the excessive repositioning in RFRRf during the hours of low demand is reduced by using a lower value for γ. However, with $\omega_z^{sup(v)}$ the RFRRf consistently provides better performance than other methods. The weighted sum approach with γ of 0.7 mainly performs better than PBD solution due to lower excessive repositioning during the hours of low customer demand. Otherwise, both solution approaches perform quite similarly during the daytime.

Figure 5.18 shows the exact percentage improvements of the RFRRf method with varying values of γ and Δs_{cell} for vehicle supply estimation using $\omega_z^{sup(v)}$. With a value of 0.7 for γ and 500 m for Δs_{cell}, the RFRRf serves 5.3% more customers and produces 3.3% percent more profit than min-distance. The performance improvement is even higher for larger regions; with Δs_{cell} of 1500m the RFRRf serves 6.7% more customers with a profit improvement of 4.8% over min-distance. RFRRf even outperforms RFRR method with γ of 0.7 and Δs_{cell} of 1500m; it serves 4.5% more customers and generates 2.5% more profit.

Overall the above results show that the RFR methods provide the most consistent AMoD performance gains even when the spatial accuracy of the forecast varies. This shows the importance of considering the ID function inside repositioning methods.

Perfect versus Imperfect Forecast

Besides the spatial accuracy of the forecast, it is also important to analyze the performance of the repositioning methods when the temporal aspect of the forecast information has errors. Any advanced forecast method used to estimate future customers should be at least better than directly
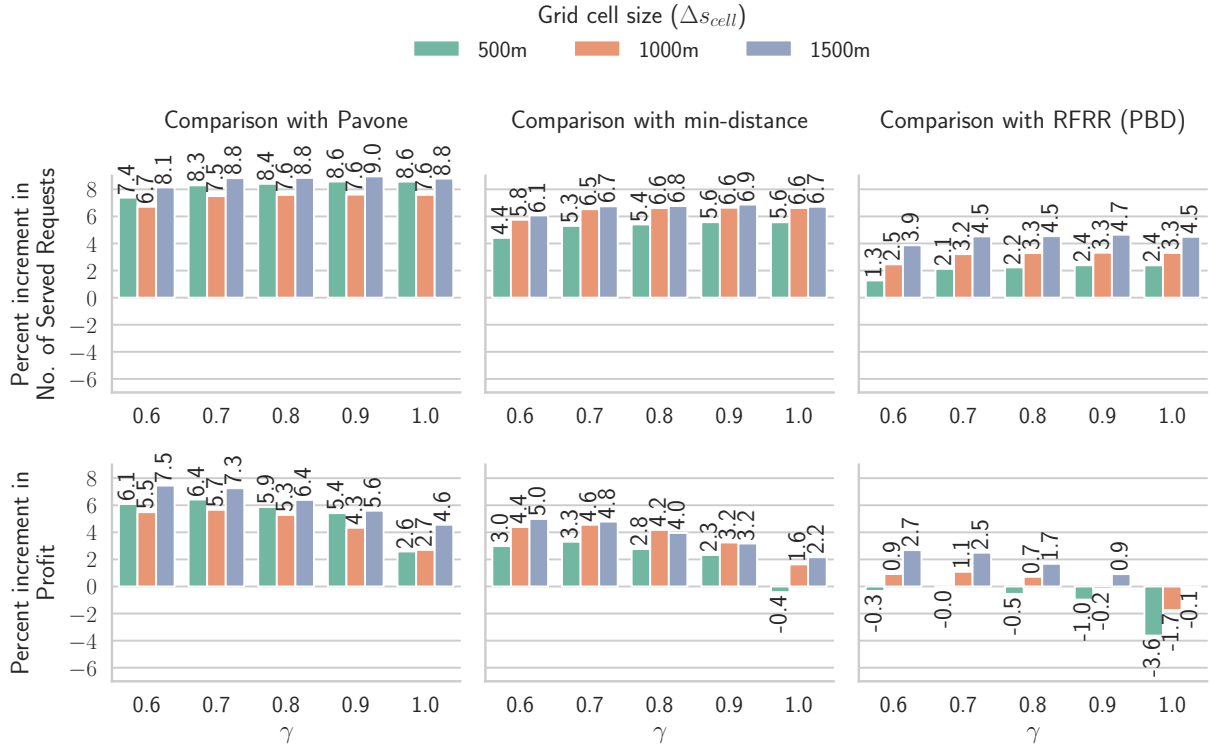
Figure 5.18.: The performance of RFRRf method with supply-side estimate using $\omega_z^{sup(v)}$ and ΔT_h of +30 minutes. By adjusting the values of γ, RFRRf outperforms other methods, especially for larger regions.

using the information from the immediate past. Therefore, this section uses the regional weights generated using the customer information of the last 30 minutes (imperfect forecast) to set a lower bound of the AMoD performance using the least accurate forecast method possible.

As shown in Figure 5.19, the general behavior of the repositioning methods observed in previous sections remains the same with a slight decrease in performances when the imperfect forecast is used. Figure 5.20 shows the relative performance differences. As Figure 5.20a shows, the combination of $\omega_z^{sup(c)}$ and the imperfect forecast shows a consistent performance drop for all methods especially for larger Δs_{cell}. A major reason for this is that $\omega_z^{sup(c)}$ uses customer destination locations as an estimate of long-term vehicle supply. The imperfect forecast causes inaccuracies in predicting the aggregated customer origins and destinations, which causes errors in not only demand forecast but also in the estimations of long-term vehicle supply. Since the min-distance and Pavone's method are more sensitive to vehicle supply estimations than RFR methods, the additional errors in vehicle supply estimation cause the most performance drop for these methods. For the min-distance and RFRR method, there is a decrease in performance with the increase of Δs_{cell}: with large Δs_{cell}, the repositioning vehicles can be distributed to a smaller number of points (due to smaller number of regions) from where they additionally have limited access to all points in the region within ΔT_{max}. The performance drop of RFRR is smaller than min-distance as it already tries to cover bigger areas due to the usage of RFs.

In contrast to the above, the performance drop of Pavone's method decreases with the increase of Δs_{cell} up to 1000 m, suggesting that equal distribution of excess vehicles makes it less susceptible to errors in the imperfect forecast for bigger regions. After 1000 m, the performance drop of Pavone's

(a) Vehicle supply estimate using $\omega_z^{sup(c)}$

(b) Vehicle supply estimate using $\omega_z^{sup(v)}$
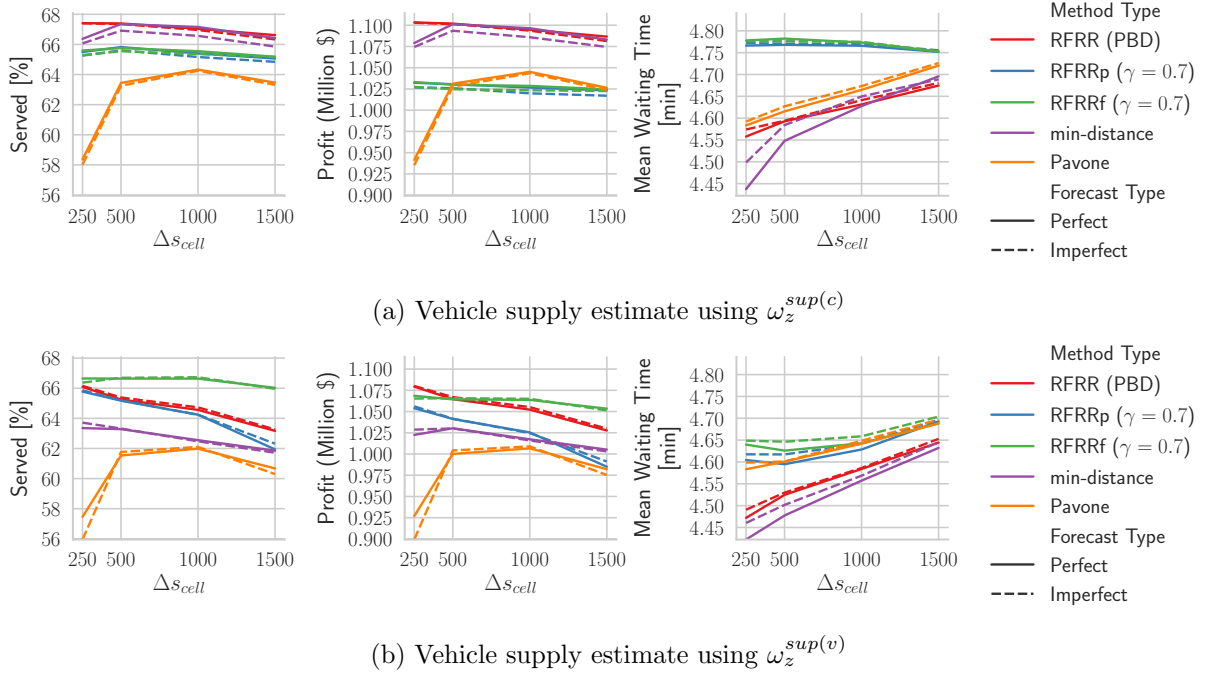
Figure 5.19.: The impact of spatial accuracy and supply estimations on the repositioning methods. The perfect and imperfect forecasts have ΔT_h of +30 and -30 minutes, respectively.

method again increases due to repositioning vehicles not being able to be distributed well. Similar to Pavone's method, the performance drop of RFRRf method also decreases with increasing Δs_{cell}. A major reason for this is that the RFRRf methods try to balance the supply-demand ID of the operation area as much as possible without any regional constraints. A large Δs_{cell} already causes the repositioning problem to have a smaller number of repositioning points; the estimation errors of imperfect forecast in some of these regions would have little impact on RFRRf method as it tries to balance multiple regions using supply-demand ID.

Contrarily to the above, the combination of $\omega_z^{sup(v)}$ and the imperfect forecast does not show a specific pattern for performance drop with increasing Δs_{cell}; some of the methods instead perform slightly better with the imperfect forecast. However, these performance gains could be due to the dynamic nature of the AMoD simulation and not necessarily due to the repositioning method itself. One reason for such behavior could be that the imperfect forecast only affects the estimate of customer demand in $\omega_z^{sup(v)}$, while the supply-side estimate remains the same for the perfect and imperfect forecast. This error in demand estimates may introduce some randomness to the simulation and cause the overall AMoD performance to provide different results.

Overall, the results in this section show that the performance differences are relatively low when an imperfect forecast is used with the repositioning methods. This indicates that even temporally less accurate forecast methods can significantly improve AMoD services by employing repositioning methods.

(a) Vehicle supply estimate using $\omega_z^{sup(c)}$.

(b) Vehicle supply estimate via $\omega_z^{sup(v)}$.
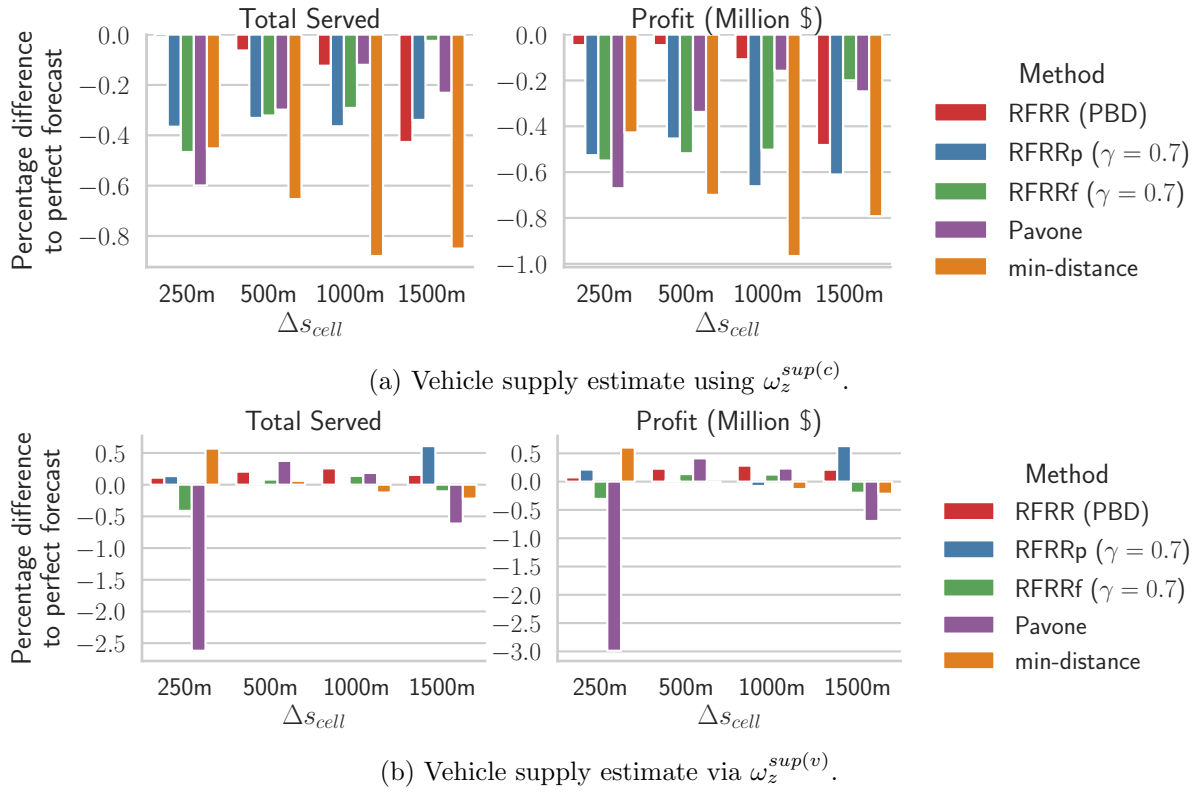
Figure 5.20.: Relative difference of using imperfect forecast relative to perfect forecast.

Forecast Horizon ΔT_h

The previous sections assumed a constant forecast horizon ΔT_h of 30 minutes. However, many times, due to various statistical reasons, the forecast algorithms provide higher accuracy when predicting for larger time windows. Therefore, this section studies the impacts of varying ΔT_h on the performance of repositioning methods. As shown in Figure 5.21, a fundamental characteristic to note is that the impact of ΔT_h on the performance of each repositioning method heavily depends on the forecast information used for vehicle supply estimation.

As shown in Figure 5.21a, the impact of increasing ΔT_h on min-distance, RFRR and Pavone's method is comparatively small with vehicle supply estimate using $\omega_z^{sup(c)}$. This shows that a rough estimation of vehicle supply using customer destination locations is sufficient for gaining AMoD performance using these algorithms. Interestingly, a short-term forecast ($\Delta T_h = 15$ minutes) deteriorates the performance of the min-distance method. A major reason for this is that even after repositioning decisions are made, idle vehicles take some time to reach the destination locations. This may cause the repositioned vehicles to be unable to pick up some expected customers within ΔT_{max}. With higher ΔT_h, the regional weights are applicable for extended periods, and the repositioned vehicles can serve a higher percentage of expected customers. This is also indicated by a lower value of W_{mean} for higher ΔT_h; it shows that a higher number of vehicles were able to pick up the customers near the repositioning points, causing a lower W_{mean}. However, if the ΔT_h is too high, the repositioned vehicle might have to wait a lot for the next customer as the customer may appear at any time in the large window ΔT_h, leading to a deteriorated performance. It is also noted that RFRR method performs better than min-distance for ΔT_h of 15 minutes; the usage of

(a) Vehicle supply estimate via $\omega_z^{sup(c)}$.

(b) Vehicle supply estimate via $\omega_z^{sup(v)}$.
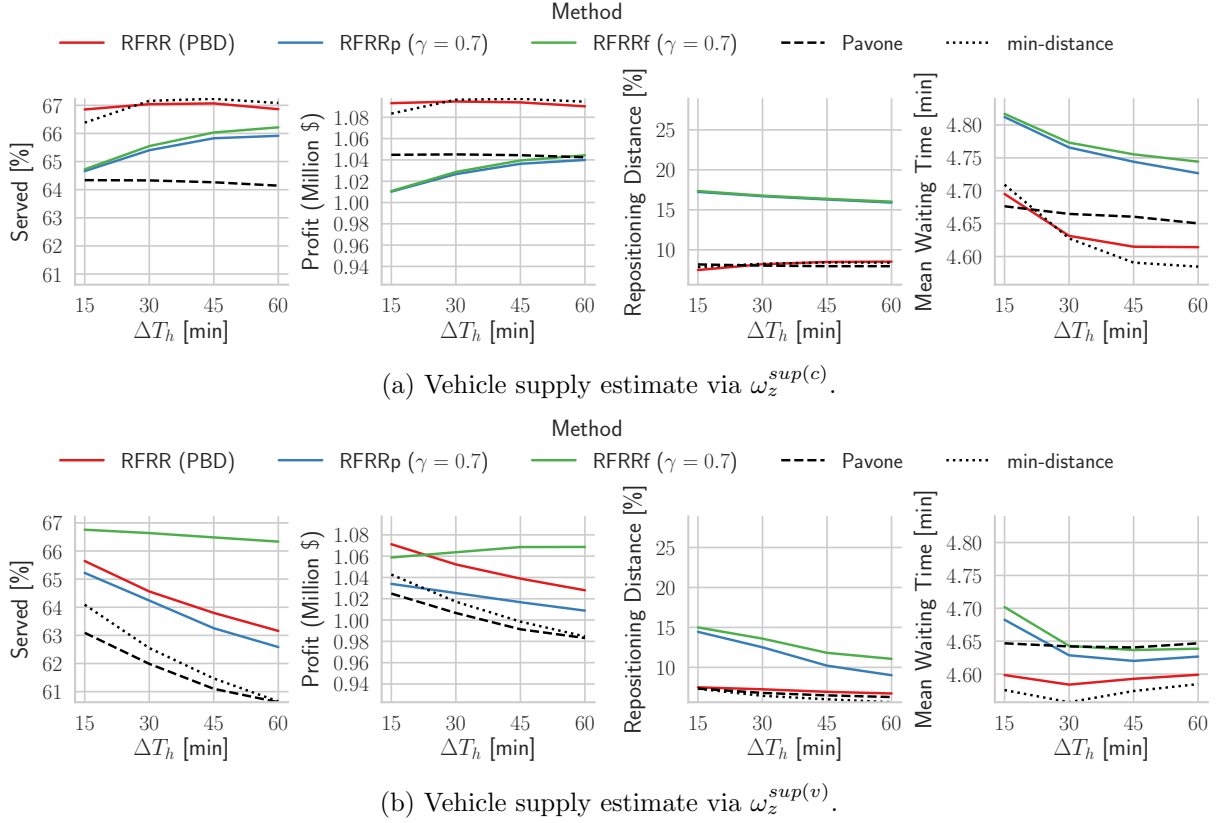
Figure 5.21.: Impact of forecast horizon ΔT_h on repositioning method with $VCO_{enroute}$ and Δs_{cell} of 1000 m.

RFs makes the RFRR to cover multiple regions with each repositioning decision, and thus, even if the repositioning vehicles take some time to reach the destination, they still have a higher chance to serve customers from multiple regions.

Interestingly, in stark contrast to the above, the performance of RFRRf and RFRRp methods improve as ΔT_h increases. Since the RFRRf is quite sensitive to the accuracy of vehicle supply-demand estimate, the $\omega_z^{sup(c)}$ for larger ΔT_h can more accurately estimate supply-demand imbalance that is applicable for more extended periods. As a result, repositioning decisions of RFRRf remain valid for longer periods, and since RFRRf already tries to minimize the ID as much as possible (due to no regional constraints for repositioning), the AMoD fleet can cover a large area for longer period. This decreases the necessity for frequent repositioning. Overall, this leads to a higher performance of RFRRf for larger values of ΔT_h.

Figure 5.21b shows the variation of ΔT_h when $\omega_z^{sup(v)}$ is used for vehicle supply estimate. In contrast to Figure 5.21a, here all methods except RFRRf show a significant drop in performance as the value of ΔT_h increases. A major reason for this is the strict regional constraints for repositioning idle vehicles; idle vehicles can only be repositioned out of the regions of vehicle oversupply. As ΔT_h increases, the magnitude of regional demand also increases (due to a larger number of customers being aggregated), and since $\omega_z^{sup(v)}$ only uses the current status of vehicles to calculate the vehicle supply, a higher number of regions appear undersupplied to the repositioning methods. This leads to fewer vehicles being repositioned and a deteriorated overall performance for these methods. In contrast, since RFRRf does not have regional constraints for repositioning vehicles, the apparent

(a) VCO_{idle}

(b) $\text{VCO}_{enroute}$
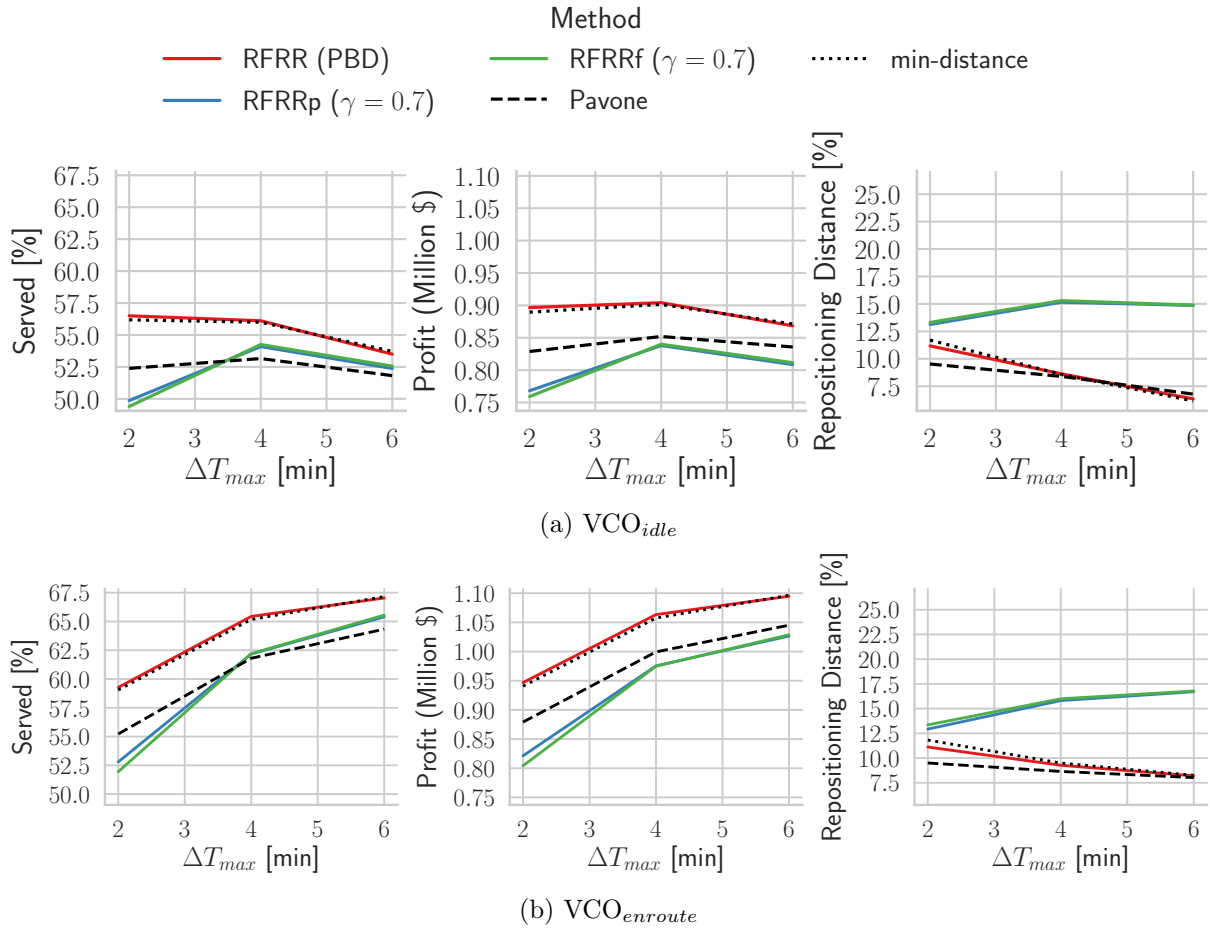
Figure 5.22.: The Impact of maximum waiting time ΔT_{max} on the performance of repositioning methods.

regional under-supply due to large ΔT_h does not affect its repositioning decisions. On the contrary, using the information of only current vehicle supply accompanied with long-term customer demand (as explained in the previous paragraph) improves the overall performance of RFRRf method with larger ΔT_h. In fact, it outperforms all other methods for ΔT_h greater than 15 minutes. This makes RFRRf especially important when the forecast method used has a high temporal resolution and can only forecast the pickup locations of the customers.

5.6.3. Maximum Waiting Time

The maximum waiting time allowed for picking up a customer ΔT_{max} plays a significant role in the service quality of the AMoD services. This section studies the performance of repositioning methods when the value of ΔT_{max} is changed, as shown in Figure 5.22. Since VCO uses ΔT_{max} as a hard constraint during the assignment process, a larger value of ΔT_{max} would mean that the vehicles could be assigned to customers located at larger pickup distances. For the RFR methods, this would imply that RFs have larger bandwidths. Contrarily, a smaller ΔT_{max} would mean that the RFs have smaller bandwidths, and thus, the individual regions will be comparatively independent of neighboring regions.

As shown in Figure 5.22a, for the VCO$_{idle}$, ΔT_{max} needs to be within a certain range to get the most performance gain from repositioning methods. Since the VCO$_{idle}$ only assigns idle vehicles to customers, a small ΔT_{max} limits the possible customers that can be picked up in the vicinity of the available vehicles. These limited customers may not necessarily produce the highest profit; due to smaller ΔT_{max}, the VCO has a smaller choice of customers to be assigned to vehicles which may force VCO to assign vehicles to any customer available even if they produce very small variable fare. In comparison, with a large ΔT_{max}, the VCO would have a large choice of vehicles for customer assignments, from which it would assign customers that produce the highest AMoD profit. This can be observed by the slight increase in the profit of RFRR and min-distance methods as ΔT_{max} is increased from 2 minutes to 4 minutes, even though the 2 minutes variant has higher $S_\%$ than 4 minutes. Interestingly, with ΔT_{max} higher than 4 minutes, the $S_\%$ and the profit is decreased for VCO$_{idle}$; since the current chapter uses a myopic VCO that only considers batch optimization without re-optimization of assigned customers, a larger ΔT_{max} would mean that the VCO would be forced to also assign vehicles to customers in the current batch that have higher pickup distances. This means that the vehicles will remain busier and will not be available for repositioning or to serve some other customers in the immediate future with shorter pickup distances. Thus, for VCO$_{idle}$, choosing the right ΔT_{max} is essential for getting the best system performance.

In contrast to the above, the performance of VCO$_{enroute}$ increases with the value of ΔT_{max}. Following the same argument as above, a larger ΔT_{max} allows the AMoD service to assign customers from larger pickup distances. However, contrary to the VCO$_{idle}$, the VCO$_{enroute}$ batches have more vehicles available due to consideration of enroute vehicles. This increases profit and $S_\%$. However, as described in section 4.3.1, this put more customers into the schedule of individual vehicles, leading to a significantly higher W_{mean} for VCO$_{enroute}$ than VCO$_{idle}$. With a higher ΔT_{max}, the number of possible vehicles assigned to individual customers is also higher due to the larger pickup distance allowed from vehicle availability locations. Thus, the VCO$_{enroute}$ can pack an even higher number of customers into AMoD fleet schedules, leading to a higher $S_\%$ and profit.

Here it should also be noted that even though VCO$_{enroute}$ leads to higher W_{mean}, nevertheless, this higher W_{mean} can be adjusted by using a smaller ΔT_{max}. Ultimately, the individual AMoD operator will have to decide if the increase in $S_\%$ and the profit is worth the decrease in the service quality (W_{mean}) for VCO$_{enroute}$. Additionally, VCO$_{enroute}$ comes with the additional risk of unplanned delays caused by the dropping off of the last customer and the changing traffic conditions. In these cases, the AMoD operator may consider compensating the customer monetarily for any additional delays caused by assigning an enroute vehicle.

5.6.4. Repositioning Frequency

The repositioning period ΔT_r can significantly impact the overall AMoD performance. A lower ΔT_r would mean that the idle vehicles are repositioned more frequently and vice versa for higher ΔT_r. Even after the repositioning methods are applied, the supply-demand imbalance can keep reemerging due to the differences in the pattern of customer origins and destinations. Therefore, the FC must regularly reposition the idle vehicles to avoid AMoD performance drop. This phenomenon is observed in Figure 5.23, where the performance of almost all repositioning methods decreases by increasing ΔT_r. Interestingly, the RFRR, Pavone's, and the min-distance method do not show a significant drop in repositioning VKT even with increased ΔT_r; however, the $S_\%$ is significantly dropped. This shows that the frequent repositioning of idle vehicles is really important for maintaining a high AMoD performance with the repositioning methods.

The RFRRp and RFRRf show a slightly different behavior than other repositioning methods. Since the RFRRp and RFRRf are prone to excessive repositioning, the increased ΔT_r significantly
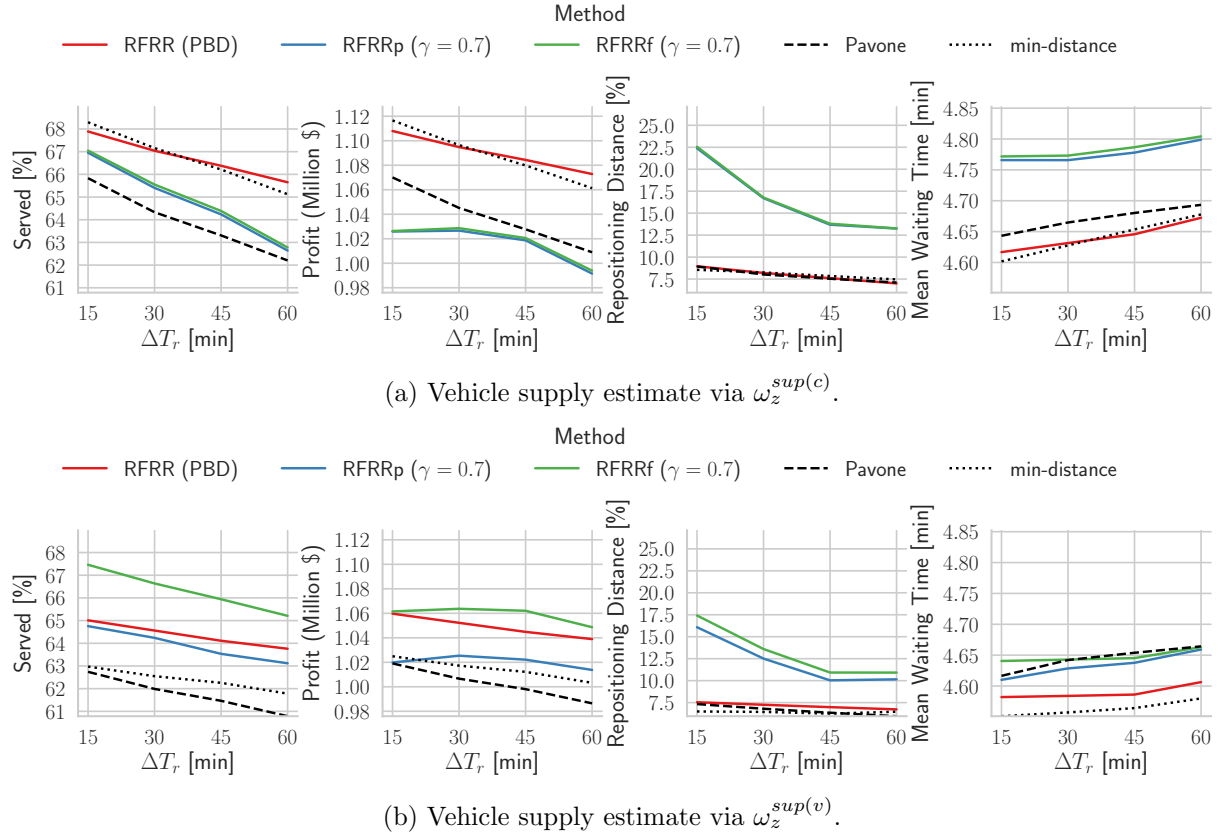
(a) Vehicle supply estimate via $\omega_z^{sup(c)}$.

(b) Vehicle supply estimate via $\omega_z^{sup(v)}$.

Figure 5.23.: The impact of repositioning period ΔT_r. The simulations used $VCO_{enroute}$ and ΔT_h of 30 minutes.

reduces the repositioning VKT. Similar to other methods, this also decreases $S_\%$. However, since the repositioning costs also reduced due to lower VKT, the overall impact on the AMoD profit is comparatively lower than other repositioning methods.

5.6.5. Reachability Frequency Bandwidth

While all the previous sections used $AP_{(90,90)}$ for the RF bandwidths, this section studies the influence of using other RF bandwidths. For simplicity, the section mainly focuses on the RFRR and RFRRf methods as shown in Figure 5.24. Since RFRR and RFRRf works the best with vehicle estimates using $\omega_z^{sup(c)}$ and $\omega_z^{sup(v)}$, respectively, Figure 5.24 only shows AMoD performance with with these combinations. The RF bandwidths are varied by changing the parameters of the adaptive percentile (AP) method, described in section 4.2.4. Additionally, Figure 5.24 also shows changes in performance when instead of ID based relation matrix, a unity matrix is used for the $\boldsymbol{A}$ matrix in Eq. 5.10. With the unity matrix, the RFR methods consider the individual regional weights for repositioning independently.

First, for the RFRR method, there is not a significant performance difference between an AP bandwidths and the unity matrix as shown in Figure 5.24a. A major reason is that the RFRR only repositions idle vehicles from surplus to deficient regions, which hinders utilizing the full potential of ID based repositioning. Additionally, due to the realistic network travel times, sending more vehicles directly to the regions of higher demand (due to the squared objective function and unity matrix in

(a) RFRR (PBD solution) and vehicle estimate using $\omega_z^{sup(c)}$

(b) RFRRf (with γ of 0.7) and vehicle estimate using $\omega_z^{sup(v)}$
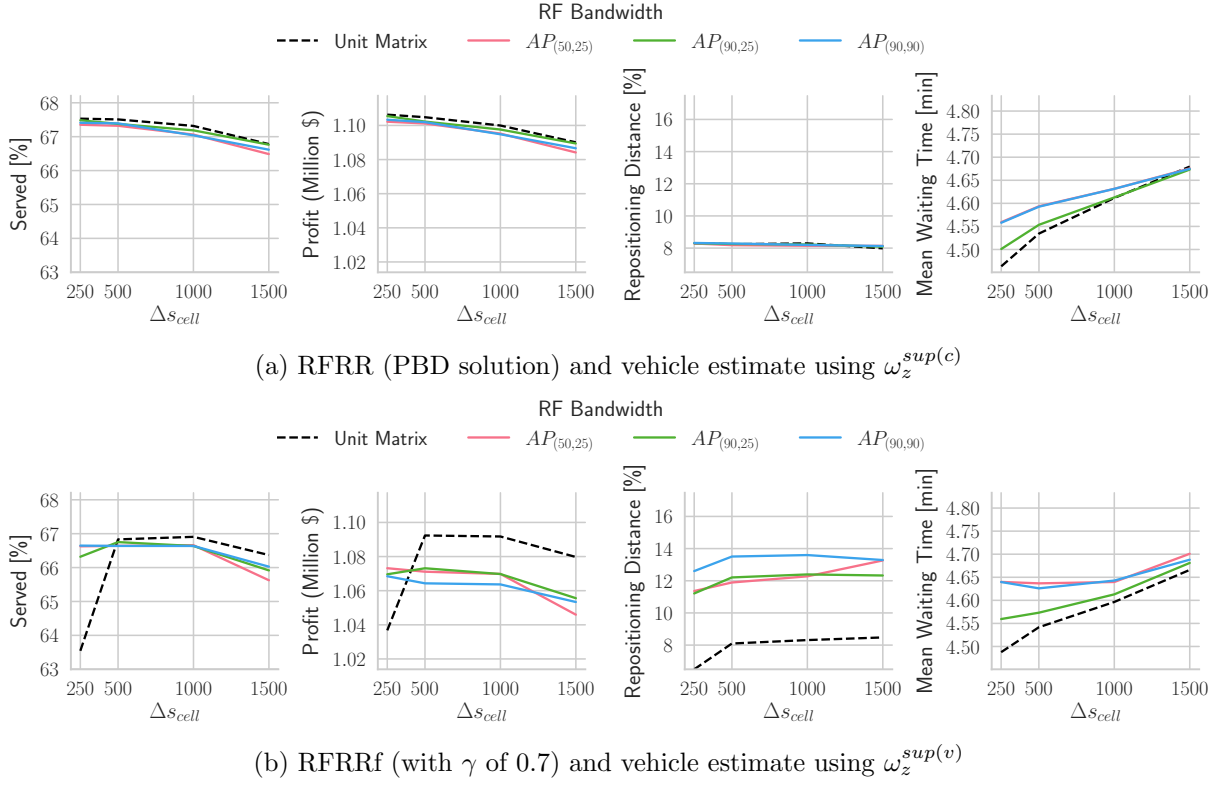
Figure 5.24.: Impact of various RF bandwidths on repositioning methods with $\text{VCO}_{enroute}$ and ΔT_h of +30 minutes.

Eq. 5.13) makes the AMoD vehicles to be more accessible to customers.

In contrast to the above, the RFRRf method shows a quite different behavior than RFRR method as shown in Figure 5.24b. First, the combination of only using current vehicle locations for the vehicle supply estimate $(\omega_z^{sup(v)})$ and minimizing the expected ID provided the best performance with the RFRRf method. However, the benefits of using ID are even more visible for the smaller Δs_{cell} in Figure 5.24b: 1) Since the repositioning methods use the region-based estimation of ID, smaller Δs_{cell} provides better inter-regional relationships due to higher spatial resolution. 2) The smaller Δs_{cell} provides better distribution of vehicles as the vehicles have a higher number of locations (the regional centers) where they can be relocated. Thus, for the Δs_{cell} of 250 m, the RFRRf has higher $S_\%$ and AMoD profit than the unity matrix. However, the unity matrix performs better for the higher Δs_{cell} due to the higher inaccuracy in ID estimation and a limited number of vehicle repositioning points.

5.6.6. Fleet Sizing and Pricing Structure

Figure 5.25 shows the AMoD performance with increasing fleet sizes. For all fleet sizes, the first observation is that all repositioning methods have significantly higher $S_\%$ and profit than the scenario without repositioning. Additionally, the mean waiting time W_{mean} is decreased due to the higher availability of AMoD vehicles. An increase in the fleet size also causes an increase in the repositioning VKT for all methods; however, the mean repositioning VKT per vehicle remains almost similar for RFRR, min-distance, and Pavone's method. This means the repositioning VKT is scaled almost
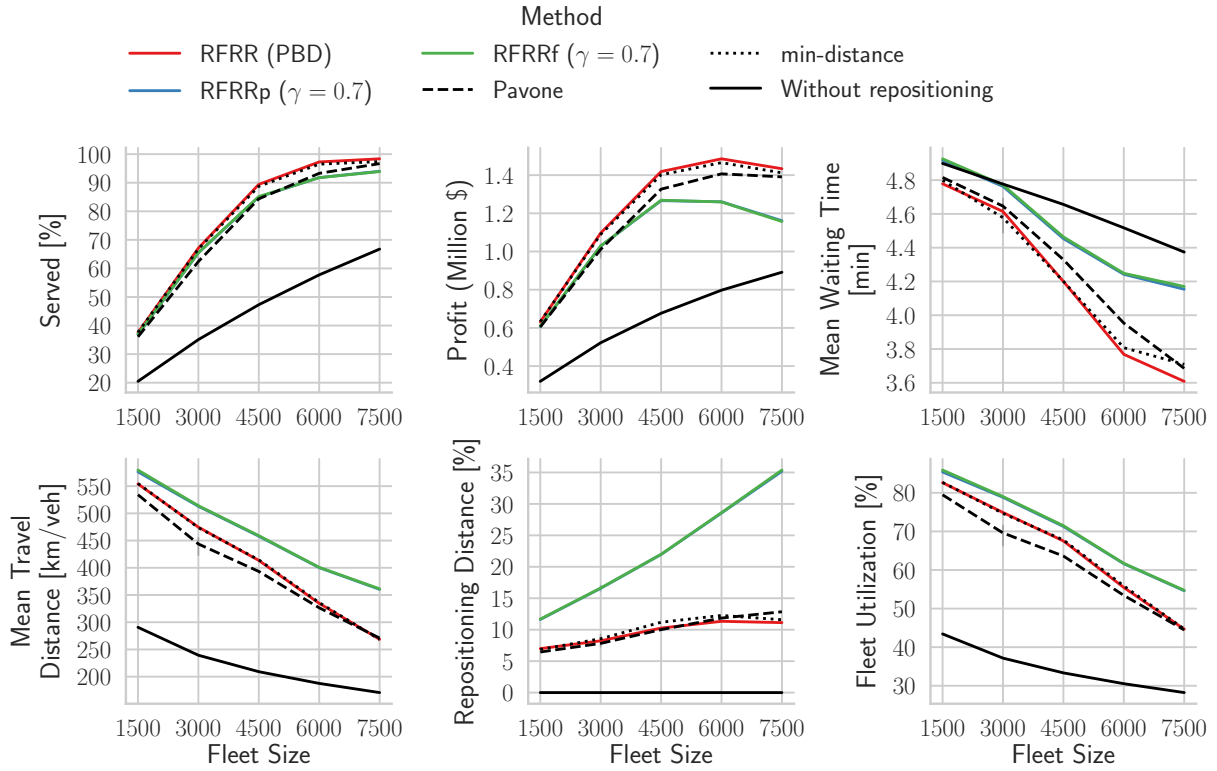
Figure 5.25.: The of impact fleet sizing on the repositioning methods using $\text{VCO}_{enroute}$ and $\omega_z^{sup(c)}$.

linearly with the additional fleet. In contrast, the RFRRp and RFRRf show an increase in the repositioning VKT per vehicle with increasing fleet size. This is also in line with the behavior observed in Figure 5.8 for the static instances; a high number of vehicles makes the RFRRp and RFRRf reposition a significantly higher number of vehicles for marginal increment in ID objective. For the RFRR method, similar to its behavior on static instances, a very high number of vehicles causes a decrease in the repositioning VKT as observed for a fleet size of 6000 and 7500.

Figure 5.26 compares the performance of VCO types and the vehicle supply estimates. Even with a high repositioning VKT of RFRRf method, the performance of RFRRf is significantly improved for all fleet sizes when $\omega_z^{sup(v)}$ used; it even outperforms RFRR method up to a fleet of size 4500. Another important observation is that after a certain fleet size, both VCO types produce similar AMoD profit with repositioning; however, the VCO_{idle} produces significantly lower W_{mean} than the $\text{VCO}_{enroute}$ for a similar AMoD profit.

In addition to the above, it is also observed that an increase in the AMoD fleet only produces additional profit up to a certain limit. Most customers can already be served at this point, and the additional fleet only increases the maintenance cost causing a decrease in the overall profit. Furthermore, the increase in AMoD profit is not linearly related to the AMoD fleet size; rather, the marginal increase in profit decreases after certain AMoD fleet sizes: the increase in profit from 4500 to 6000 vehicles is significantly smaller than when increasing the vehicles from 3000 to 4500. A major reason is that the additional fleet cannot serve an equivalently higher customer demand and remain underutilized, as indicated by $U_\%$ in Figure 5.27. This could be either due to insufficient customer demand or the stochastic nature of the overall simulation.
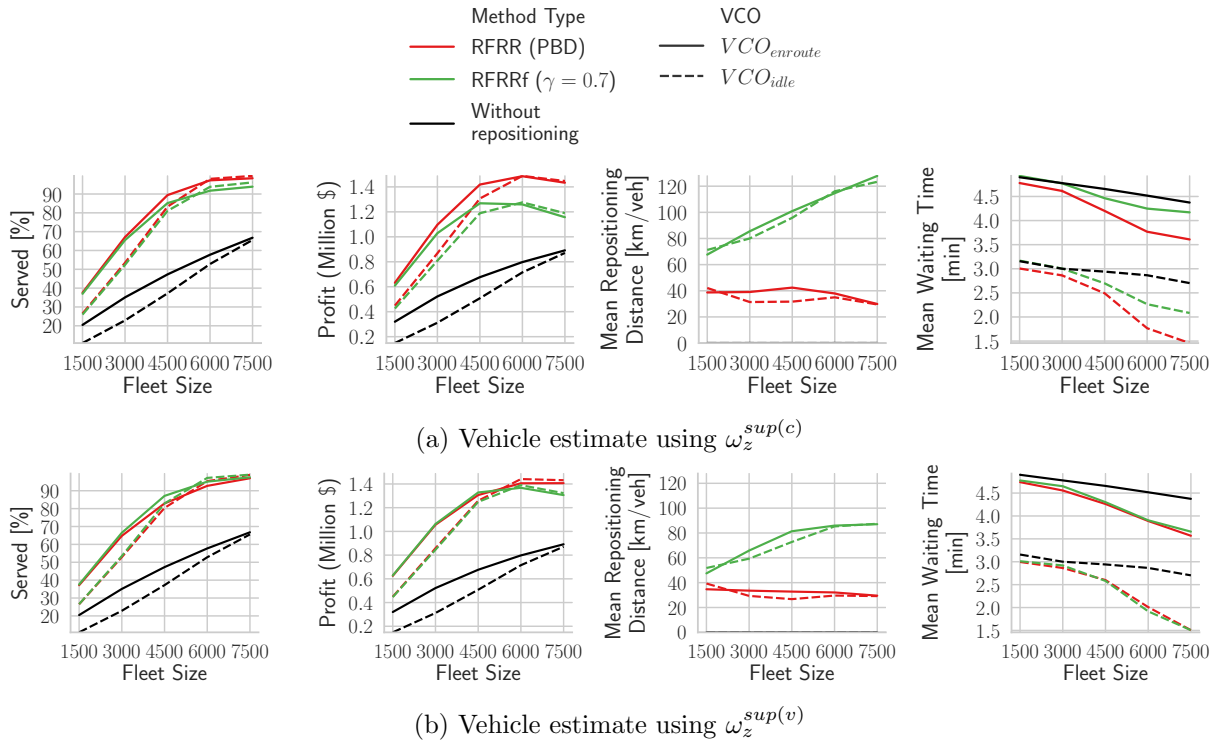
Figure 5.26.: The comparison of VCO types for different fleet sizes and vehicle supply estimates.

Figure 5.28 shows the exact AMoD performance improvement values for each repositioning method. The most important observations are as follows:

- The RFRR method provides the best AMoD performance in combination with $\omega_z^{sup(c)}$ and $VCO_{enroute}$. Compared to the scenario without repositioning, on average, it serves 32.5% more customers with a relative increase of around 92.4% in AMoD profit. It serves almost 3% more customers and produces 4.8% more profit relative to Pavone's method. Even though an increase in AMoD profit over Pavone's method may appear small, in reality, it is significantly high; for example, for the two days of simulation and a fleet of 4500 vehicles, the RFRR method produces a profit of almost $90 thousand USD more than Pavone's method, which would mean a yearly saving of almost $16.4 million.

- Although the repositioning methods produce the highest AMoD profit with $VCO_{enroute}$, the AMoD providers in future may still prefer using the VCO_{idle} for several reasons. First, the $VCO_{enroute}$ causes a significantly higher customer waiting time which is easily visible by comparing the W_{mean} of VCO_{idle} and $VCO_{enroute}$ in Figure 5.28. Second, it assigns a high proportion of customers to enroute vehicles, which, depending on the traffic state, has a high potential of causing additional pickup delays. The RFRR method also provides a significant improvement for VCO_{idle}; on average, in comparison to VCO_{idle} scenario without repositioning, it serves 34.3% more customers with a relative profit increment of 142.2%. In comparison to Pavone's method, it serves 1.54% more customers with a relative profit increment of 3.6%.

- As showed before in Figure 5.26, the performance gaps between the VCO_{idle} and $VCO_{enroute}$ are more prominent for small fleet sizes. Numerically, with RFRR method and a fleet of
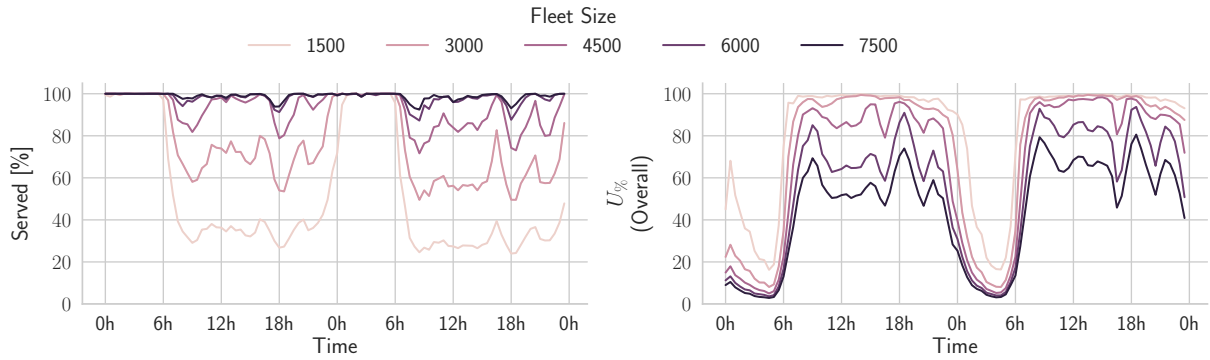
Figure 5.27.: The temporal fleet utilization of RFRR method using $VCO_{enroute}$, $\omega_z^{sup(c)}$ and varying fleet sizes.

(a) VCO_{idle}

(b) $VCO_{enroute}$
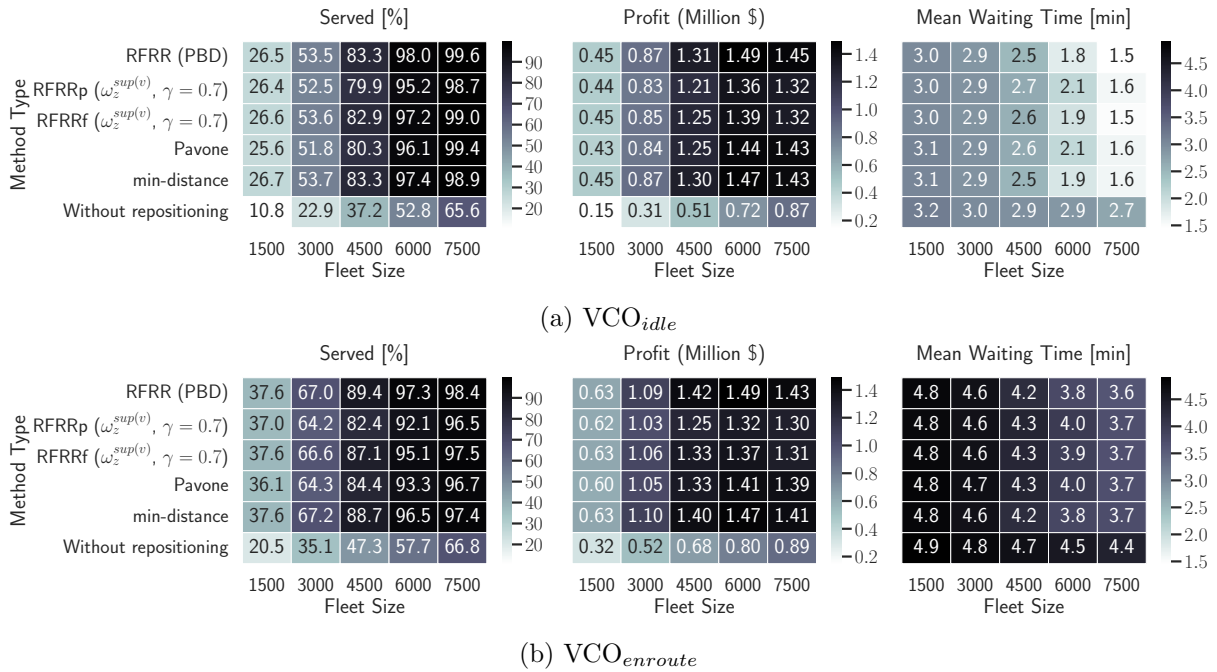
Figure 5.28.: The table of performance improvement values for different VCO types and fleet sizes. Since RFRRp and RFRRp perform the best with $\omega_z^{sup(v)}$, the table shows their values with $\omega_z^{sup(v)}$. The rest of the methods used $\omega_z^{sup(c)}$ for vehicle supply estimation.
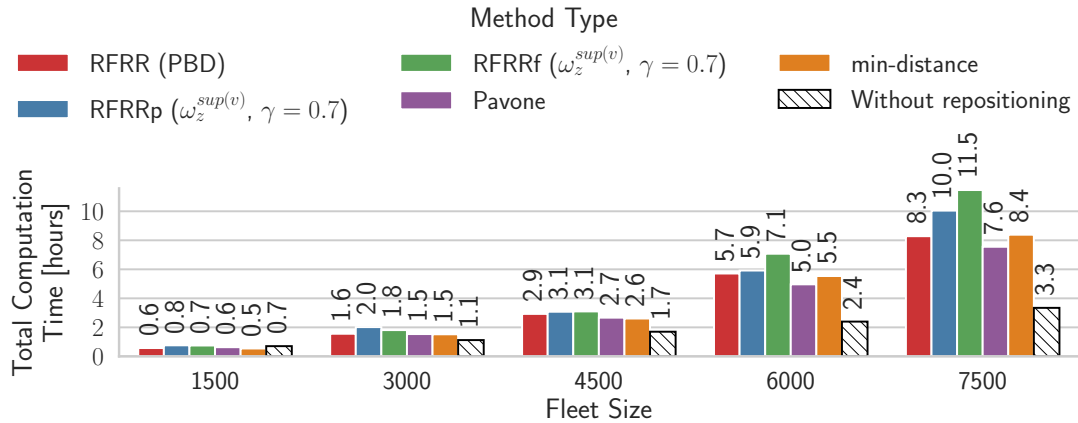
(a) Whole Simulation.
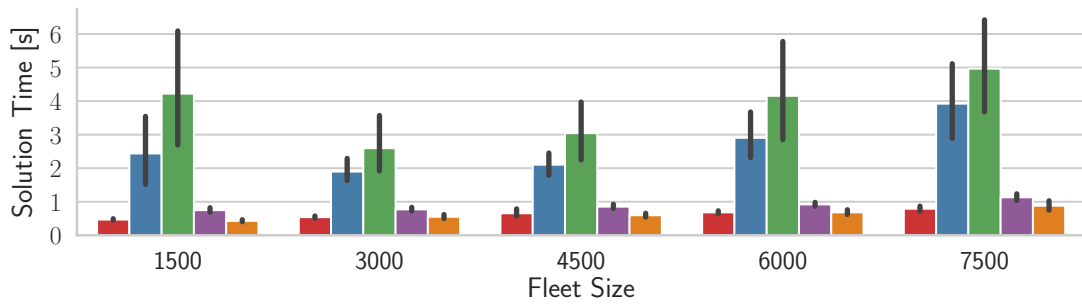
(b) Solution time of individual repositioning problem instances, including the time taken to prepare the problem. The error bars show 95% confidence interval.

Figure 5.29.: The computation time taken for two days of AMoD simulation using $VCO_{enroute}$.

1500 vehicles, the $VCO_{enroute}$ serves 11.1% more customers (40% profit increment) than VCO_{idle} which decreases to 6.1% more customers (8.4% profit increment) with a fleet of 4500 vehicles. With an even higher fleet size, the performance gain reverses and VCO_{idle} serves more customers and generates more AMoD profit than $VCO_{enroute}$ along with a significantly reduced W_{mean}.

Figure 5.29 compares the computation time taken for the two days of AMoD simulation. As shown in Figure 5.29a, the simulation of a large fleet takes significantly more time than a smaller fleet. However, the increase in computation time is due to the movement of additional vehicles and the associated processing required for simulation — not due to the time taken to solve the repositioning problem. On average, all the methods take a significantly small time to solve the repositioning, as shown in Figure 5.29b. For some problem instances, the RFRRp and RFRRf take significantly high solution time, so a time limit of 30 seconds is used with these two methods. However, this time limit is only occasionally reached by both methods. Nevertheless, the computation time for all the methods is within a practical range for all fleet sizes.

Finally, Figure 5.30 shows the changes in the AMoD profit when a different price structure is used. Since the fixed base fare charged per customer forms a major part of the price structure, the section only varies the base fare to see its impact on the AMoD profit. Figure 5.30a shows that without repositioning, the AMoD service would face loss if a base fare is not charged. A larger fleet serves more customers, which increases revenue from variable fares; however, the increased maintenance

(a) Without repositioning

(b) RFRR method
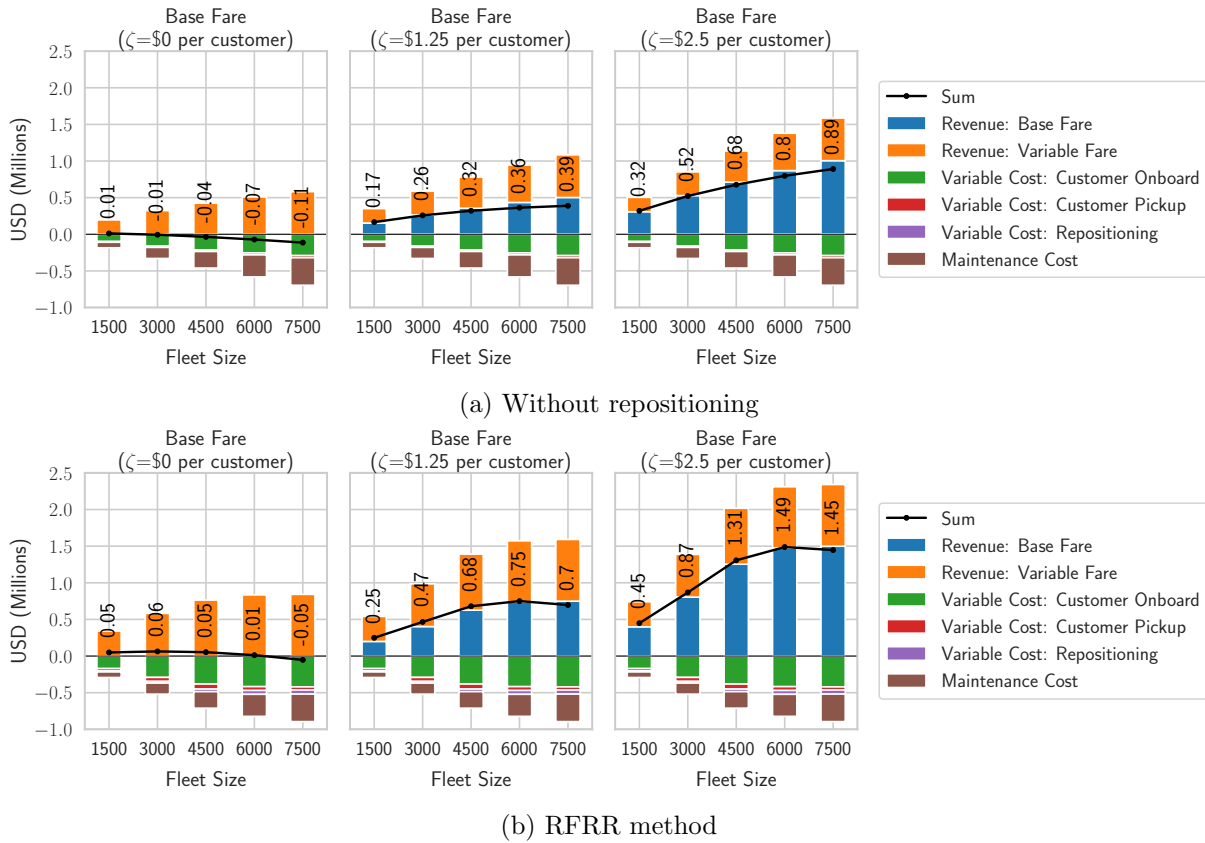
Figure 5.30.: The AMoD profit components for $VCO_{enroute}$ and different values of the base fare. The annotated numbers show the overall AMoD profit.

cost leads to an overall loss without the base fare. Even with RFRR repositioning, the AMoD service would hardly make any profit without charging a base fare as shown in Figure 5.30b. Furthermore, the repositioning and the pickup journeys contribute a comparatively small proportion to the overall costs. Here it should also be noted that this analysis does not include penalty cost for unserved customers by the AMoD service; a smaller fleet leads to smaller $S_\%$ and the customers that are rejected by the AMoD service may cause negative popularity. Thus, many studies also include a penalty term for the lost opportunity cost for unserved customers.

5.7. Conclusion

The varying pattern of the origin and destination locations of customers leads to an accumulation of AMoD fleet. It causes vehicle supply-demand imbalance in different regions of the operation area. Without human drivers, the central FC must regularly reposition the idle vehicle to demand-intensive regions to reduce this imbalance and increase the system's efficiency. This chapter introduced four new methods for this purpose. The chapter introduced several repositioning methods. The first method, called the min-distance method, tries to reposition as many vehicles as possible with minimum repositioning distance while aiming to fulfill regional demands. The other three methods (RFRR, RFRRp and RFRRf — grouped under RFR methods) utilize a region-based approximation to ID to reposition idle vehicles in a way that minimizes the long-term ID. In contrast to the previous

chapter, where ID was only used as a spatial metric to describe service quality in the form of a heat map, this chapter actively used ID to improve the performance of AMoD services.

The newly developed methods were tested on static problem instances and in an agent-based AMoD simulation. The developed methods showed a significant improvement over the benchmark method. The chapter also performed the sensitivity analysis of the most important parameters. Overall, the RFRR method produced the best performance, followed by the min-distance method.

The current chapter also showed that the repositioning frequency needs to be high to gain the best performance out of the repositioning methods. The next chapter will focus on how the vehicle supply-demand imbalance could be actively reduced inside VCO such that the need for the regular repositioning of idle vehicles is minimized.

Chapter 6.

Proactive Assignment Strategy

The previous chapter introduced reachability function based repositioning (RFR) methods for the AMoD services. The main goal of the presented methods was to reposition idle vehicles to demand-intensive regions such that a higher proportion of customers remain on the positive side of imbalance density (ID) metric introduced in chapter 4. This would mean that a higher number of customers are within reachable range of the AMoD vehicle, ultimately leading to improved AMoD performance.

The current chapter targets the same objective from another perspective; instead of regularly repositioning idle vehicles to demand-intensive regions, it modifies the VCO to minimize the long-term supply-demand imbalance. Thus, it introduces a proactive assignment method where a portion of the short-term monetary profit is compromised to reduce the long-term supply-demand imbalance. The chapter combines the presented proactive assignment approach with the RFR repositioning, specifically the RFRR method. Together with the efficient repositioning of RFRR method and the reduced repositioning requirement of the proactive assignment, they provide the best AMoD performance.

The chapter is structured as follows. Section 6.1 discusses the primary motivation behind developing a new VCO approach. Section 6.2 presents the mathematical formulation and the solution approach followed for the proactive assignment. The performance of the proactive assignment is evaluated in two steps. First, section 6.3 presents the simulation results when the proactive assignment is used alone without explicit repositioning of idle vehicles. This section also compares the performance against the RFRR performances of the previous chapter where non-proactive VCO is used. Later, section6.4 shows the performance improvement when the RFRR method is used together with the proactive approach. Section 6.3 and section6.4 also offer the sensitivity analysis of the most relevant simulation parameters. Finally, section 6.5 concludes the chapter.

6.1. Motivation

The previous chapter showed a significant AMoD performance improvement using RFR and min-distance methods. In general, repositioning idle vehicles keeps the AMoD fleet within reachable distances of customers and improves the overall performance. As shown in Figure 6.1, the repositioning frequency plays a major role in getting the most out of any repositioning method. For example, by changing the ΔT_r from 15 minutes to 60 minutes, the percentage of served customers ($S_\%$) of RFRR method drops from 67.9% to 65.6% and the AMoD profit drops from \$1.11 million to \$1.07 million — the drop is even higher for other repositioning methods. While the previous chapter showed that the AMoD performance could be significantly improved by taking better repositioning decisions, the requirement of high repositioning frequency shows that the positive impacts of repositioning decisions do not last long; the FC must frequently reposition idle vehicles to maintain a balance between AMoD supply and demand. This also contributes towards the empty VKT of AMoD fleet to perform repositioning tasks; for example, the repositioning constitutes almost 9%
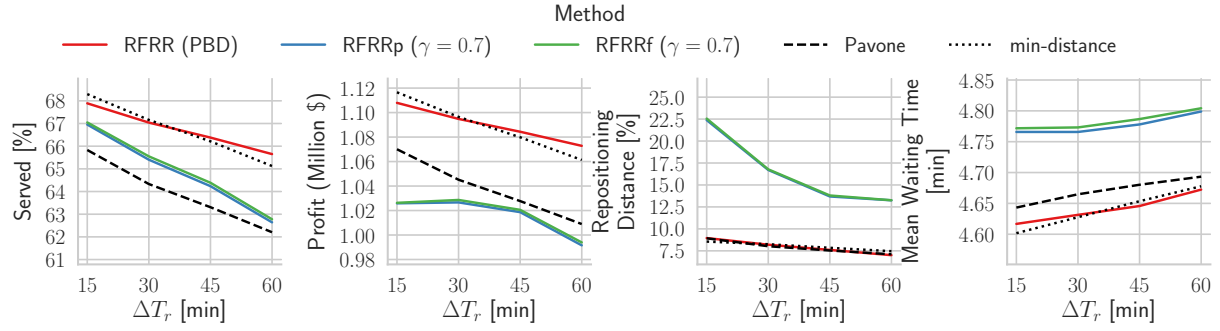
(a) Vehicle supply estimate via $\omega_z^{sup(c)}$.
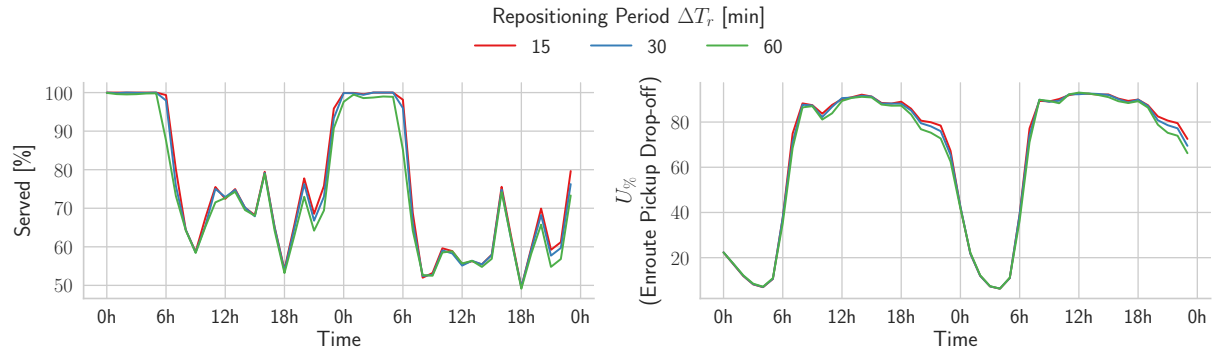
(b) Hourly performance of RFRR method

Figure 6.1.: The impact of repositioning period ΔT_r.

of the total VKT traveled using RFRR method with ΔT_r of 15 minutes. Besides the chances of getting stuck in congestion during repositioning, these additional VKT may also contribute to the congestion, especially in urban areas. Therefore, the current chapter asks how the requirement of explicit repositioning of idle vehicles — and consequently the repositioning VKT — can be minimized without significantly compromising the AMoD performance.

Fundamentally, the need for repositioning occurs because of supply-demand imbalance arising due to the pattern of customer origins and destinations, as described in section 5.1. Hence, the above question could be answered on two levels: first, on the city level, the authorities could design the cities in a way that minimizes the disparity of customer origin and destination patterns, and second, the AMoD operator optimizes its operation in a way that reduces the requirement of repositioning. Since most cities where the AMoD services will most probably be launched are well developed, manipulating the customer origins and destinations is complex and beyond the scope of the current dissertation. Instead, the present chapter focuses on the second approach.

Besides the repositioning of idle vehicles, the primary location in the AMoD operation cycle where the supply-demand imbalance could be considered is the VCO. Figure 6.2 shows the core concept followed in the chapter. Typically, a simple profit-oriented VCO leads to regional imbalances between AMoD supply and demand. The repositioning of idle vehicles periodically rectifies this imbalance; however, the imbalances reemerge after some time. In contrast, the chapter argues that the occurrence of supply-demand imbalance can be reduced by considering the long-term supply-demand imbalance in the VCO decisions (proactive assignment). This, in turn, also reduces the requirement of repositioning, improving the AMoD performance.
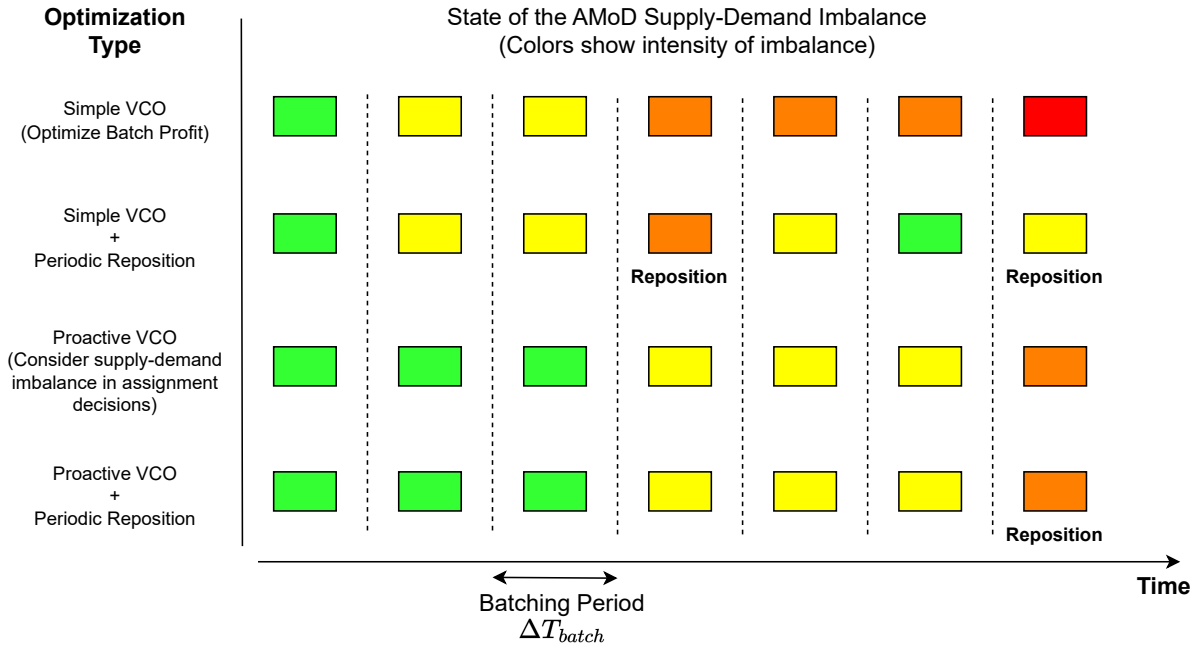
Figure 6.2.: A rough comparison of the supply-demand imbalance occurring from simple and proactive assignment strategies.

6.2. Methodology

The section presents the methodology followed to develop the proactive assignment method. Section 6.2.1 first presents a general version of the proactive assignment. The formulation assumes an already existing VCO, and thus, could be used with any VCO method. Similar to the repositioning methods, proactive assignment also requires a forecast of future AMoD system state. Therefore, section 6.2.2 presents the methods used for this purpose. Additionally, it describes a formulation to link the decision variables of VCO with the long-term supply-demand imbalance using ID function. Finally, section 6.2.3 formulates the proactive assignment approach for the VCO methods used in the dissertation.

6.2.1. General Proactive Assignment Method

This section formulates a general proactive assignment method for the VCO. The main advantage of the formulation is that it complements the already existing VCO and can be combined with any VCO method — not just the ones used in this dissertation. The general functioning of the FC remains the same as described in section 3.1.3: the dynamic aspect of the AMoD services is solved by grouping the customers into VCO batches, followed by forming and solving a bipartite matching problem for assigning vehicles to customers. The section uses the same terminologies used in section 3.1.3 to describe the proactive assignment method. Additionally, the dissertation refers to the proactive assignment approach as proactive VCO to distinguish it from simple VCO of chapter 3.

Let set R_b^t represent the customer requests in a single VCO batch at time t. Each customer $r \in R_b^t$ has an associated pickup location (p_r), drop-off location (d_r), and the time t_r when the customer requests a ride. t_r also represents the earliest possible time a vehicle could pick the customer up. The VCO assigns vehicles to customers out of a set of vehicles $V^t \subseteq V$. Each $v \in V^t$ can start its

journey to pick up the customer at time t_v from location o_v. Let $y_{vr} \in \{0,1\}^{|V^t| \times |R_b^t|}$ represent the binary decision variable for the assignment of $v \in V^t$ to $r \in R_b^t$ and $f_{vco}(\boldsymbol{y})$ represent the objective function of the chosen VCO method. With the given VCO method, the dissertation modifies it into a multi-objective formulation with $f_{ib}(\boldsymbol{y})$ as a function to estimate the long-term impacts of $\boldsymbol{y}$ on supply-demand imbalance. Thus, the formulation is given as follows:

$$\max_{\boldsymbol{y}} \quad f_{vco}(\boldsymbol{y}), -f_{ib}(\boldsymbol{y}) \tag{6.1a}$$

$$\text{s.t.} \quad \sum_{v \in V^t} y_{vr} \leq 1 \qquad \forall r \in R_b^t \tag{6.1b}$$

$$\sum_{r \in R_b^t} y_{vr} \leq 1 \qquad \forall v \in V^t \tag{6.1c}$$

$$g_i(\boldsymbol{y}) \leq 1 \qquad \forall i \in \{1, 2, ..., m\} \tag{6.1d}$$

where $g_i(\boldsymbol{y})$ are any additional constraints used by the VCO method. Eq. 6.1b and Eq. 6.1c ensure that only one vehicle is assigned to a request. Some VCO methods, instead of directly assigning a vehicle to a customer request, make possible vehicle plans for each vehicle, and then assign plans to each vehicle to maximize the objective function. The above proactive assignment method would still be valid for such VCO by simply changing the binary variable to map from V^t to vehicle plans instead of R_b^t.

The aim of the multi-objective optimization in Eq. 6.1 is that by consistently considering the system imbalance while assigning the vehicles to customers, the overall vehicle supply-demand balance could be maintained or improved in small steps. A possible downside of such an approach could be the unequal treatment of customer requests; some customers with destinations in demand-intensive regions of the operational area may be implicitly preferred over other customers. This may also lead to a high number of vehicles only serving in the demand-intensive regions, contributing to additional traffic in the region. On the bright side, such an approach would improve the system efficiency of a smaller fleet to serve many customers. Nevertheless, the suggested proactive assignment should be combined with a dynamic pricing scheme that ensures that the overall AMoD system remains equitable and provides enough incentives to the customers depending on VCO decisions. However, the frequency and magnitude of destination-based customer preference heavily depends on how the multi-objective problem in Eq. 6.1 is solved and how much the solution differs from the original optimal solution of the chosen VCO.

Since the dissertation focuses on the algorithmic aspect of the vehicle assignment process that maximizes the monetary profit of the MSP, developing a dynamic pricing scheme for system equity is beyond the scope of the dissertation. Instead, the dissertation shows the potential improvement in system efficiency and MSP profit when a proactive assignment method is used. This can serve as a foundation to incorporate the technique in an overall scheme that also considers the equal treatment of the customers.

6.2.2. Estimating the Long-Term System Imbalance using Imbalance Density function

The previous section introduced a general scheme for incorporating proactive assignment in a VCO method. The estimator function $f_{ib}(\boldsymbol{y})$ is Eq. 6.1a can be implemented in multiple ways. The dissertation proposes an estimator motivated by the region-based forecast in section 5.2.2 and RFR methods.

Let t be the simulation time the VCO is solved. The method first estimates the differences of supply and demand for all regions in Z, given by the vector $\boldsymbol{\eta} \in \mathcal{Z}^{|Z|}$. Similar to section 5.2.2, the current regional imbalance η_z for $z \in Z$ consists of parts for vehicle supply and demand: $\eta_z = \eta_z^{sup} - \eta_z^{dem}$, where η_z^{sup} and η_z^{dem} are the weights for regional supply and demand, respectively.

Let R_z^+ and R_z^- be the forecast of customer destination and origins, respectively, calculated using the perfect or imperfect forecast procedure (section 5.2.2) for time-window $[t, t + \Delta T_{ph}]$. ΔT_{ph} is the forecast period used for proactive assignment. The demand weights η_z^{dem} are simply set to the corresponding forecast of customer origins $\eta_z^{dem} = |R_z^-|$. As for the vehicle supply weights η_z^{sup}, two formulations are used as described below.

The first formulation uses the current vehicles available in each zone: let $V_z \subseteq V^t$ represent all vehicles that have their availability point o_v in z and availability time t_v within the time-window $[t, t + \Delta T_{max}]$, then the vehicle supply estimate is given as:

$$\eta_z^{sup(v)} = |V_z| \tag{6.2}$$

The second formulation also includes estimating vehicles that may end up in z by serving a customer during the forecast window $[t, t + \Delta T_{ph}]$. Similar to section 5.2.2, these future vehicle supply are estimated using customer destinations R_z^+, and thus, the vehicle supply is given as:

$$\eta_z^{sup(c)} = |V_z| + |R_z^+| \tag{6.3}$$

With the above given, the changes in the regional weights can be calculated by counting the number of vehicles entering and leaving zones with each VCO decision. Let $R_z \subseteq R_b^t$ represent all customers with their destinations in z and $R_{z-} = R_b^t - R_z$ represent all customers with destinations in the rest of the zones. Then an estimate of long-term change in the regional weights is described by a vector of integer variables $\boldsymbol{\delta\eta} \in \mathbb{R}^{|\mathcal{Z}|}$, given as:

$$
\begin{aligned}
\delta\eta_z(\boldsymbol{y}) &= \delta\eta_z^{sup}(\boldsymbol{y}) - \delta\eta_z^{dem}(\boldsymbol{y}) && \forall z \in Z \\
&= \underbrace{\sum_{r \in R_z} \sum_{v \in V^t} y_{vr}}_{\substack{\text{No. of vehicles} \\ \text{entering zone}}} - \underbrace{\sum_{v \in V_z} \sum_{r \in R_{z-}} y_{vr}}_{\substack{\text{No. of vehicles} \\ \text{leaving zone}}} && \forall z \in Z
\end{aligned}
\tag{6.4}
$$

For simplicity, the function $\delta\eta_z(\boldsymbol{y})$ will be henceforth written as $\delta\eta_z$. Note that any vehicle that starts and ends in the same region will have zero contribution to $\delta\eta_z$. Next, an RF based formulation for the $f_{ib}(\boldsymbol{y})$ can be defined following a similar procedure as in section 5.4.2, given as:

$$
\begin{aligned}
f_{ib}(\boldsymbol{y}) &= f_{ib}(\boldsymbol{\delta\eta}) \\
&= F\left(\frac{3}{\pi} \sum_{z \in Z} \left(\eta_z^{sup} + \delta\eta_z^{sup} \right) \frac{k(x, x_z, h_z)}{h_z^2}, \quad \frac{3}{\pi} \sum_{z \in Z} (\eta_z^{dem} - \delta\eta_z^{dem}) \frac{k(x, x_z, h_z)}{h_z^2} \right)
\end{aligned}
\tag{6.5}
$$

where the first term and second term in Eq. 6.5 represent the long-term impact on supply and demand density with each VCO decision. The function F is a parameter function to measure the difference between the two densities. The dissertation uses the integral of squared deviations for this purpose. Thus, following the same steps as described for RFR methods (refer to section 5.4.2 for detailed steps), Eq. 6.5 can be written as:

$$f_{ib}(\boldsymbol{\delta\eta}) = a^2 (2\boldsymbol{\eta}^T + \boldsymbol{\delta\eta}^T) \boldsymbol{A} \boldsymbol{\delta\eta} + C \tag{6.6}$$

where $a = \frac{3}{\pi}$ and $C = a^2\boldsymbol{\eta}^T\boldsymbol{A}\boldsymbol{\eta}$. $\boldsymbol{A}$ is calculated using the mid-point rule, as described in section 5.4.2. Since C is a constant, it can be ignored for VCO. The formulation in Eq. 6.6 uses an RD based relation for $f_{ib}(\boldsymbol{\delta\eta})$.

It should also be noted that as the bandwidth of all regions approaches zero (i.e. the influence of neighboring regions is not taken into account), $\boldsymbol{A}$ becomes a unity matrix and Eq. 6.6 reduces to:

$$f_{ib}(\boldsymbol{\delta\eta}) = a^2 \sum_{z\in Z}(\eta_z + \delta\eta_z)^2 \tag{6.7}$$

In this formulation, the proactive VCO treats each region as independent of other regions and prefers assigning vehicles to customers going in the destination zone of the highest supply-demand imbalance.

6.2.3. Proactive Assignment for the Vehicle Control Optimization

The previous sections described the general proactive assignment method and how the long-term supply-demand imbalance can be estimated based on the decisions of a general VCO. This section modifies the formulation to be explicitly applied to the VCO used in the dissertation. The term "proactive" is added in front of the name of specific VCO to indicate that the proactive assignment method is used in combination with the VCO, i.e., proactive VCO$_{enroute}$ or proactive VCO$_{idle}$. The section also describes the method used for solving the multi-objective optimization problem of the proactive VCO.

First, the VCO used in the dissertation uses constraint on the maximum pickup time ΔT_{max}. However, instead of using the hard time constraint directly Eq. 6.1d inside VCO, the dissertation first creates a bipartite graph for the possible combination of vehicles and customers as described in section 3.1.3. Thus, the primary change required is to adapt Eq. 6.4 for the decision variables defined over the bipartite graph.

Consider $R_G \subseteq R_b^t$ and $V_G \subseteq V^t$ represent the customers and vehicles with least a single edge in the bipartite graph, respectively. Let $R_z \subseteq R_G$ be the customers with their destinations in $z \in Z$ and $V_z \subseteq V_G$ be the vehicles with their availability points o_v in z. Additionally, let $V_{(r)} \in V_G$ be the set of all vehicles connected to the request $r \in R_G$ in the bipartite graph, and vice versa for all requests $R_{(v)} \in R_G$ connected to v. Let u_{vr} be the binary decision variable for each of the edges in the bipartite graph, then Eq. 6.4 is modified as:

$$\delta\eta_z(\boldsymbol{u}) = \delta\eta_z^{sup}(\boldsymbol{u}) - \delta\eta_z^{dem}(\boldsymbol{u}) \qquad\qquad \forall z \in Z$$
$$= \sum_{r\in R_z}\sum_{v\in V_{(r)}}u_{vr} - \sum_{v\in V_z}\sum_{r\in R_{(v)}}u_{vr} \qquad \forall z \in Z \tag{6.8}$$

Thus, the proactive assignment method for the VCO used in the dissertation is given as:

$$\max_{\boldsymbol{u}} \quad f_{vco}(\boldsymbol{u}), f_{ib}(\boldsymbol{\delta\eta}) \tag{6.9a}$$

$$\text{s.t.} \quad \sum_{v\in V_{(r)}}u_{vr} \leq 1 \qquad \forall r \in R_G \tag{6.9b}$$

$$\sum_{r\in R_{(v)}}u_{vr} \leq 1 \qquad \forall v \in V_G \tag{6.9c}$$

$$f_{vco}(\boldsymbol{u}) = \sum_{v\in V_G}\sum_{r\in R_{(v)}}(\zeta + (f^D - c^D)d_r^{pd} - c^D d_{vr}^{op})u_{vr} \tag{6.9d}$$

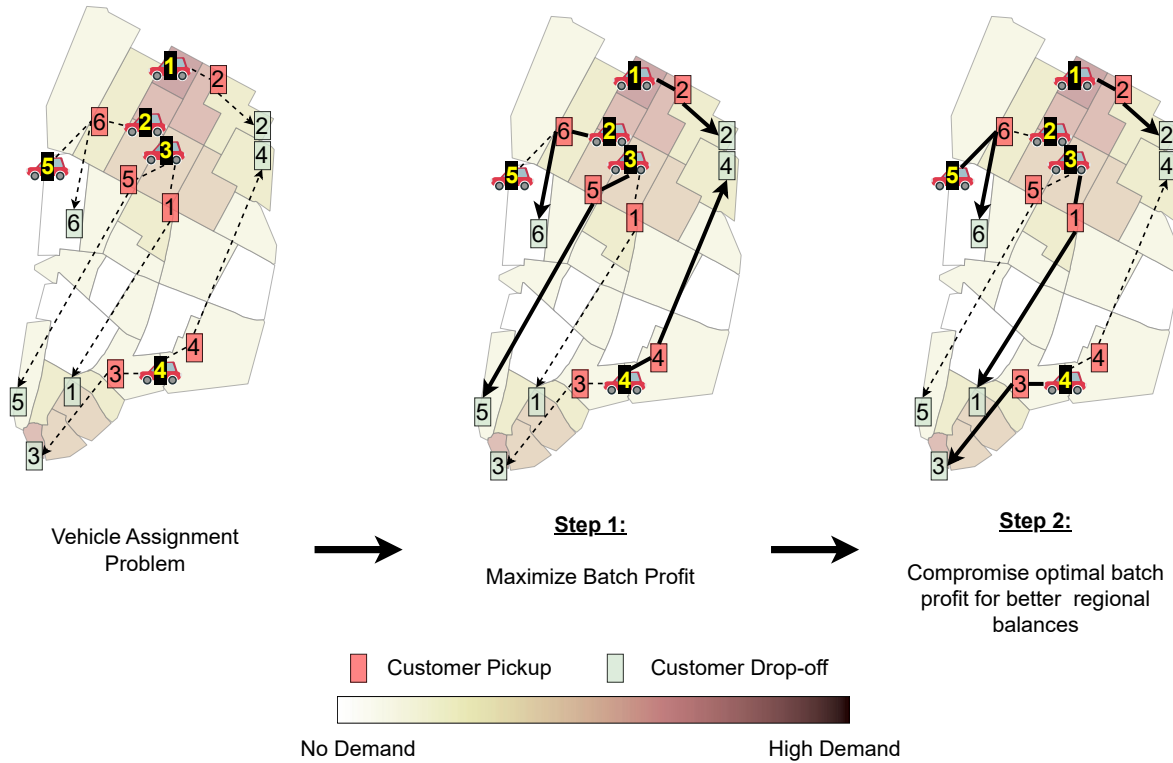$$f_{ib}(\delta\eta) = a^2(2\boldsymbol{\eta}^T + \boldsymbol{\delta\eta}^T)\boldsymbol{A}\boldsymbol{\delta\eta} \tag{6.9e}$$

Figure 6.3.: An example of proactive assignment method. The regional colors represent the magnitude of demand forecast vector $\boldsymbol{\eta}$. The dotted lines denote feasible assignments while bold lines show the assignment solution.

Many times the two objectives in Eq. 6.9a could be contradictory to each other; an assignment that reduces the long-term supply-demand imbalance might lower the immediate profit of the VCO batch. Additionally, since the original purpose for using batches with VCO is to increase the overall profit $f_{vco}(\boldsymbol{u})$ of the dynamic problem, too much focus on improving the long-term supply-demand imbalance might also lead to an overall degraded performance. Thus, the dissertation proposes a lexicographic two-step approach for solving Eq. 6.9 as follows.

First, the profit objective $f_{vco}(\boldsymbol{u})$ is maximized as a single objective, providing the optimum profit $\bar{f}$. In the second step, the long-term imbalance objective $f_{ib}(\boldsymbol{\delta\eta})$ is minimized while adding the following lower bound on the profit term:

$$f_{vco}(\boldsymbol{u}) \geq (1 - \frac{\theta}{100}) \cdot \bar{f} \tag{6.10}$$

where θ is the percentage of short-term profit an MSP is ready to compromise to minimize the long-term supply-demand imbalance. The value of θ could be either constant or dynamically controlled, e.g. as a function of recent vehicle utilization and expired requests. However, the dissertation only demonstrates the benefits of the introduced assignment concept with a constant θ. Unlike step one, the second step does not have a unimodular constraint matrix due to the additional constraint, and thus, the integer constraint $\boldsymbol{u} \in \{0,1\}^{|E|}$ cannot be relaxed for the second step.

Fig. 6.3 illustrates an example scenario in lower and midtown Manhattan areas to portray the concept. The optimal profit solution disproportionately sends vehicles to low-demand areas to maximize the profit of the current batch, i.e., vehicles 2,3, and 4 are assigned to have short pickup distances

Parameter / Method	Symbol	Section for Sensitivity Analysis	Default Value / Strategy
Base fare	ζ	—	$2.5 per customer
Distance based variable fare	f^D	—	$0.5 per km
Distance based cost	c^D	—	$0.25 per km
Fixed maintenance cost of vehicle	c^F	—	$25 per vehicle per day
Maximum allowed waiting time of customers	ΔT_{max}	6.4.2	6 minutes
The time period used for travel time scaling	ΔT_{scale}	—	30 minutes
Travel time scaling method used	—	—	ASM (section 3.3)
Fleet size	—	6.3.1, 6.4.3	3000 AVs
Vehicle assignment method	VCO	6.3.1, 6.4.3	$VCO_{enroute}$ (section 3.1.3)
Batching period	ΔT_{batch}	—	30 seconds
RF bandwidth calculation method	—	6.3.2	$AP_{90,90}$ (section 4.2.4)
The repositioning method		6.4.3	RFRR
The repositioning period	ΔT_r	6.4.1	30 minutes
The forecast horizon for repositioning	ΔT_h	—	+30 minutes (Perfect forecast)
The forecast horizon for proactive VCO	ΔT_{ph}	6.3.2	+15 minutes (Perfect forecast)
Grid cell size of regions in the operation area	Δs_{cell}	6.3.2	1000 m
The vehicle-supply estimate method for repositioning	ω_z^{sup}	—	$\omega_z^{sup(c)}$
The vehicle-supply estimate method for proactive VCO	η_z^{sup}	6.3.2	$\eta_z^{sup(v)}$

Table 6.1.: The simulation configuration used in chapter 6.

and long distances with customers on board. However, the second step partially compromises the profit in order to assign vehicles 3 and 4 to customers going to high-demand regions, i.e., to customers 1 and 3, respectively. Similarly, it assigns alternative vehicles to some customers to leave the originally assigned vehicles in high demand regions, i.e., it assigns vehicle 5 to customer 6 instead of vehicle 2. Additionally, some assignments may remain unaltered in the second step depending on the value of θ, e.g., the assignment of vehicle 1.

6.3. Experiments and Sensitivity Analysis for Proactive Assignment Method

This section performs the sensitivity analysis for the proactive VCO method. It uses the same agent-based simulation environment as used in previous chapters. The evaluation is divided into two sections. This section studies the scenario where the proactive VCO is used without repositioning. Though a comparison is made with the best-performing repositioning method in the previous chapter, i.e., RFRR; however, the repositioning configuration in this section does not use the proactive VCO
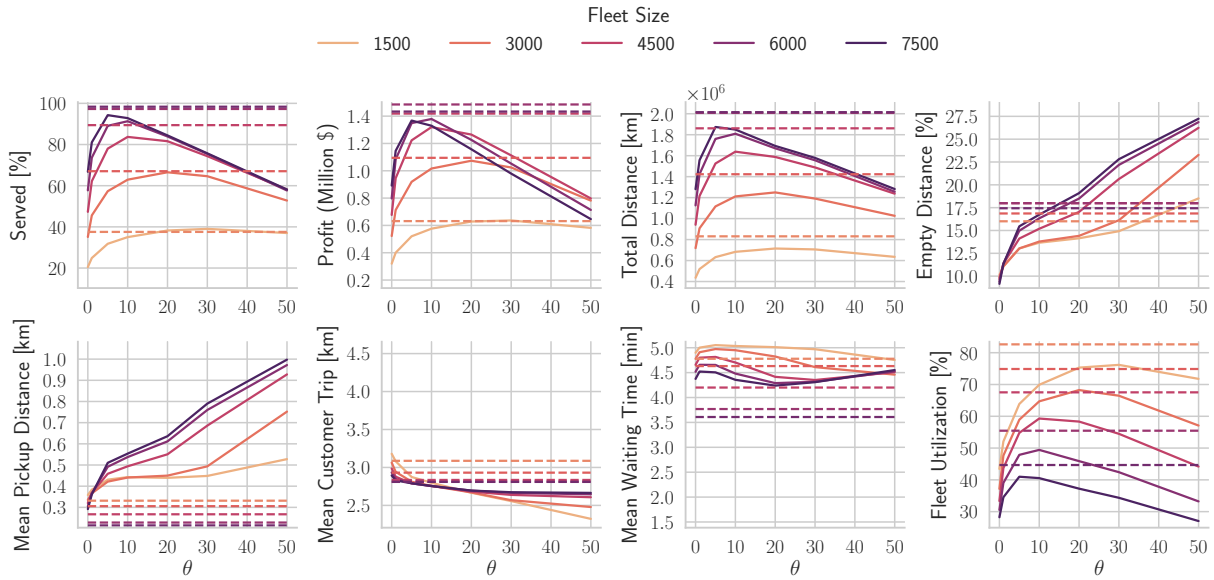
Figure 6.4.: The AMoD performance using proactive VCO$_{enroute}$ and $\eta_z^{sup(v)}$. The θ of zero corresponds to the VCO$_{enroute}$ without proactive assignment. The dotted lines represent the values of PBD solution of RFRR method under equivalent settings but without proactive assignment.

together with the RFRR method in the same simulation. The purpose is to study the impact of proactive assignments in isolation from repositioning methods. The following section will study their performances when combined in a single simulation. Both sections also present the sensitivity analysis of the most critical parameters.

Table 6.1 summarizes the default simulation configuration used in the chapter along with the detail of the specific section where a particular parameter is varied. For consistency, the base parameters remain the same as described in section 4.3. Since RFRR provided the best performance in the previous chapter, this chapter uses it as the default repositioning method.

6.3.1. The Percentage Compromise and varying Fleet Sizes

Figure 6.4 and Figure 6.5 show the performance improvement achieved by varying the θ for multiple fleet sizes. The percentage of served customers ($S_\%$) and the AMoD profit increases significantly as the value of θ increases. A particular value of θ, generally between 5 to 30%, provides the best performance for each fleet size. The method can provide significant improvements even with a slight compromise in the short-term batch profit, leading to better vehicle distribution and overall improved performance. However, a very high value of θ degrades the performance quality again, which can be explained by the higher percentage of empty distances; for higher values of θ, the proactive assignment aggressively balances the regions favoring the trips with destinations in demand-intensive areas irrespective of the empty distances traveled for picking up customers. These long trips keep the vehicles busy for longer periods with decreasing probability that the next customers could be served within the ΔT_{max} — these potential customers were the main reason for assigning customers with large pickup distances in the current VCO batch. With decreasing probability of serving these future customers, the overall performance degrades.

The above effect can also be observed in Figure 6.6 for the temporal performance; the overall
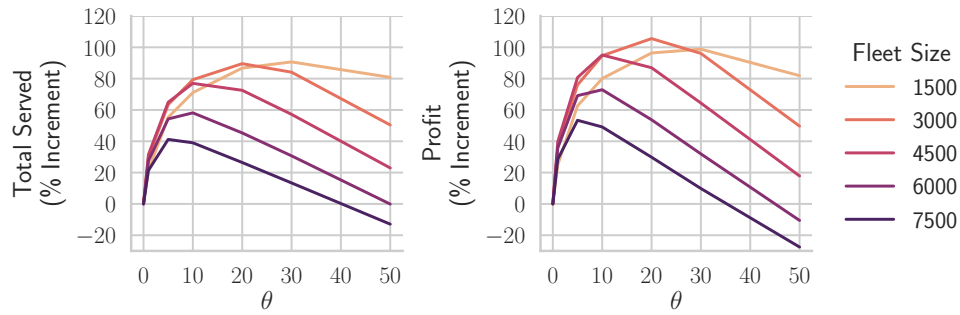
Figure 6.5.: The percentage increment of proactive $VCO_{enroute}$ over simple $VCO_{enroute}$ using varying fleet sizes. The θ of zero corresponds to the $VCO_{enroute}$ without proactive assignment.
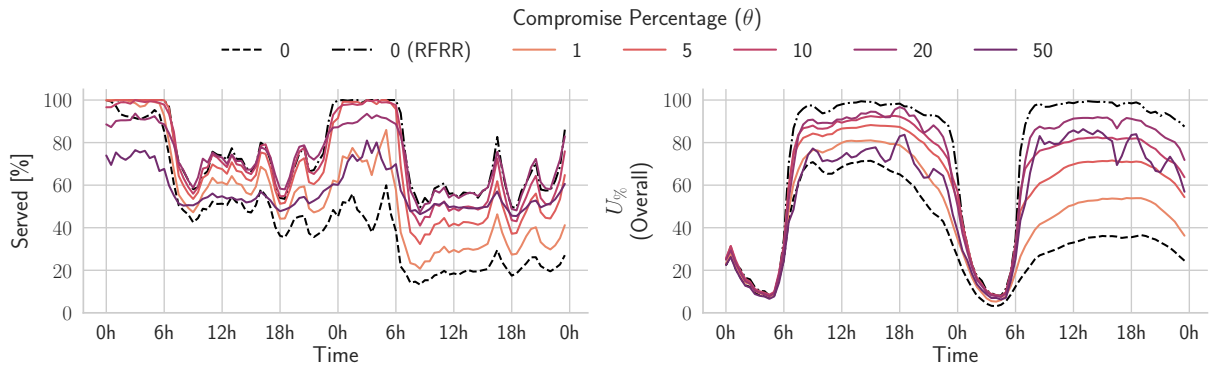
Figure 6.6.: The impact of θ on the temporal performance of proactive $VCO_{enroute}$ with a fleet of 3000 vehicles. The θ of zero corresponds to the $VCO_{enroute}$ without proactive assignment.

utilization $U_\%$ fluctuates for the higher values of θ. This shows that not all the enroute AMoD vehicles could find the next customer immediately; otherwise, there would be less variation in $U_\%$[1]. Compounding this with the fact that the system already made a higher compromise in the batch profit leads to declined performance for higher θ. This effect is more dominant for large fleet sizes. For a small fleet of 1500 vehicles, the customers are comparatively in abundance, and the vehicle is most likely able to find the next customer near the destination, reducing the negative impact of longer pickup distances. Thus, as Figure 6.5 shows, the optimal θ corresponding to the highest incremental profit shifts to lower value ranges of θ as the fleet size increases.

Figure 6.4 also shows the performance of RFRR method without proactive VCO. Using proactive $VCO_{enroute}$, the AMoD performance reaches very close to the RFRR under similar AMoD settings. As the fleet size increases, the RFRR method performs significantly better than a proactive $VCO_{enroute}$. This follows a similar reasoning as discussed for the RFRRf method (section 5.6.1) in addition to the reasons mentioned at the beginning of this section; the region-based ID used in the proactive assignment approach does not put any regional restriction on the assignment procedure based on the regional weights of the forecast. This causes the proactive method to generate a significantly high empty VKT for larger fleets when very high values of θ are used. Another major difference observed

---

[1]This fluctuation in $U_\%$ is different from the ones due to varying customer demand. The latter would also be visible in other high performing AMoD configurations like RFRR.

(a) Whole simulation.

(b) Computation time of required for solving Individual VCO batches. The error bars indicate 95% confidence interval.
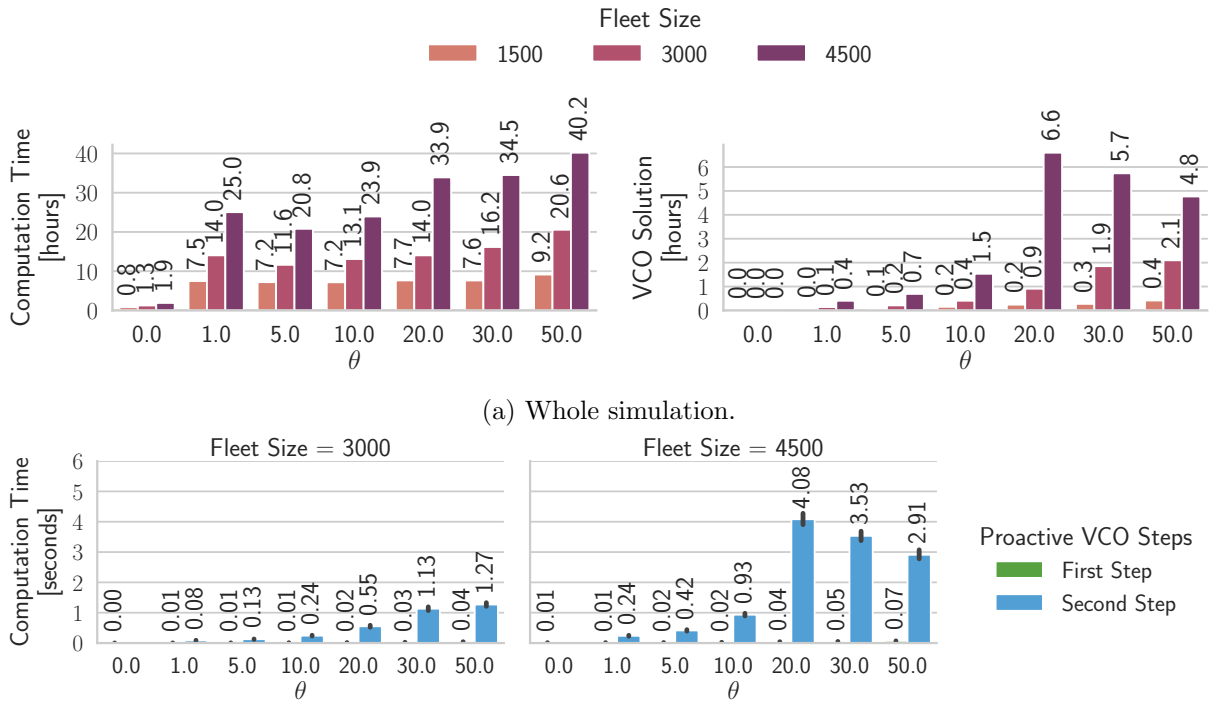
Figure 6.7.: The computation time taken by proactive VCO$_{enroute}$.

between RFRR and the proactive method is that the former does not have a particular preference on which kind of customers are served — it purely depends on the VCO objective — while the latter prefer trips that would lead to lower supply-demand imbalance, which in case of Manhattan leads to preferring shorter customer trips for high values of θ. In terms of temporal performance, Figure 6.6 shows that while the explicit repositioning consistently produces high $S_\%$ for all simulation times, the performance of proactive VCO could depend on the time of the day; for a fleet of 3000 vehicles, a θ of 20% produces the highest $S_\%$ during the daytime and a θ of 5% produces the highest $S_\%$ from 12 to 6 am (though during the night time, number of customers is comparatively low). Overall, the RFRR method performs only marginally better than the proactive method with θ of 20% and a fleet of 3000 vehicles: the RFRR method increases the AMoD profit by 109.6% and proactive VCO by 105.6% relative to simple VCO$_{enroute}$.

Figure 6.7 shows the computation of proactive VCO. The computation time of proactive VCO is significantly higher than simple VCO (i.e. when θ is zero). A major contributing factor is the simulation of increased vehicle routes due to larger fleet size and higher utilization. However, the time required for solving VCO also plays a major role in increased computation time, especially the second step of the proactive VCO, as shown in Figure 6.7b. With increased values of θ, the VCO allows higher compromise of the optimal AMoD profit, leading to an increased search space and computation time. Additionally, the formulation of proactive VCO does not have a unimodular constraint matrix due to which the integer constraint cannot be relaxed, unlike simple VCO (the first step of proactive VCO). This also generally makes the second step take longer to find the solution. It is also found that similar to the RFRRp and RFRRf methods, sometimes the second step of proactive VCO takes a significantly long time to find the optimal solution. Therefore, the dissertation uses a
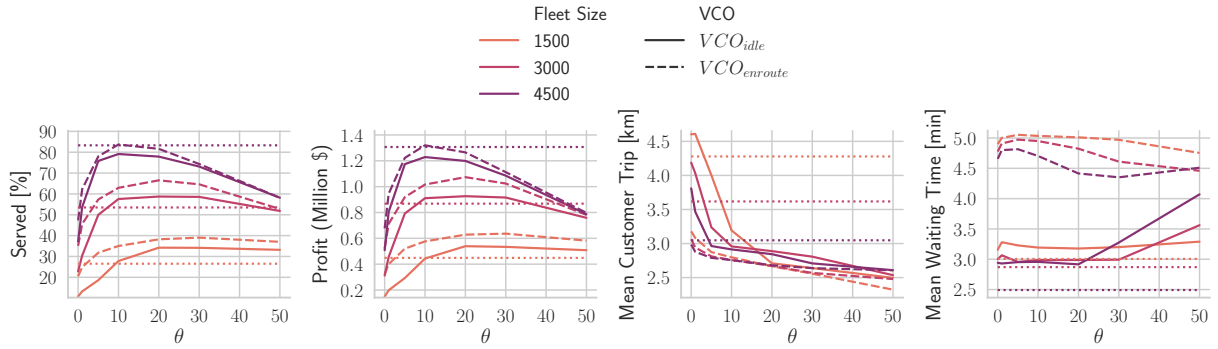
Figure 6.8.: A comparison of using proactive assignment approach with different VCO types. The horizontal dotted line represents the solution of RFRR repositioning combined with VCO_{idle} without proactive assignment.
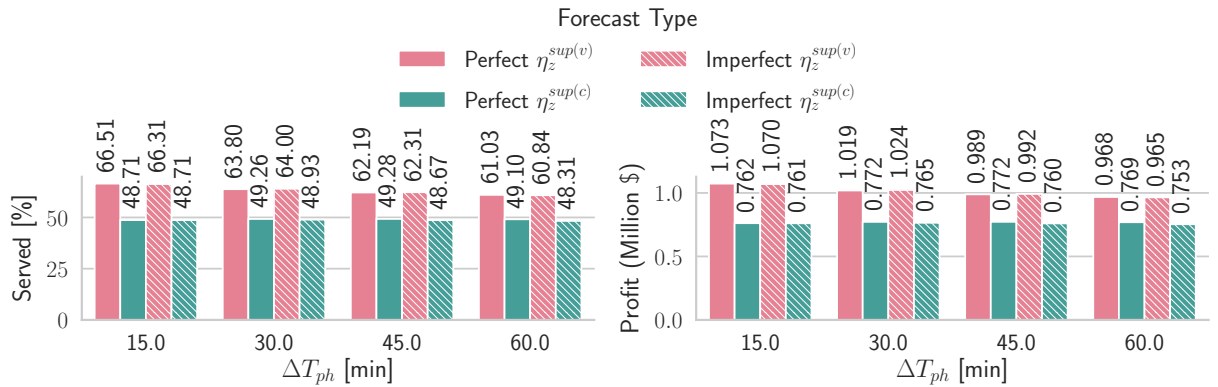
Figure 6.9.: A comparison of the temporal accuracy and type of forecast on the performance of proactive $VCO_{enroute}$.

time limit of 30 seconds for the second step; nevertheless, as visible from Figure 6.7b, the solution time is mostly well below this time limit.

Finally, Figure 6.8 compares the performance of different VCO types when used together with the proactive assignment. The proactive assignment method compliments the original VCO method and increases the method's overall performance. However, the general characteristics of the original VCO method remain similar: the $VCO_{enroute}$ produces higher $S_\%$ and AMoD profit while VCO_{idle} produces lower customer waiting times (W_{mean}). Interestingly, with moderate fleet sizes, the proactive VCO_{idle} performs even better than the combination of RFRR repositioning and VCO_{idle}. The RFRR method still performs better for larger fleet sizes than proactive VCO_{idle}.

6.3.2. Forecast Types and Reachability Function Bandwidth

Figure 6.9 compares the impact of forecast horizon ΔT_{ph} and the vehicle-supply estimate on the performance of the proactive assignment method. First, the usage of perfect or imperfect forecast information (section 6.2.2) does not have a significant impact on the performance of the proactive VCO. The perfect forecast produces slightly higher $S_\%$ and profit due to the usage of actual customers in the future in contrast to the imperfect forecast where the customer data of the last ΔT_{ph} period is used.
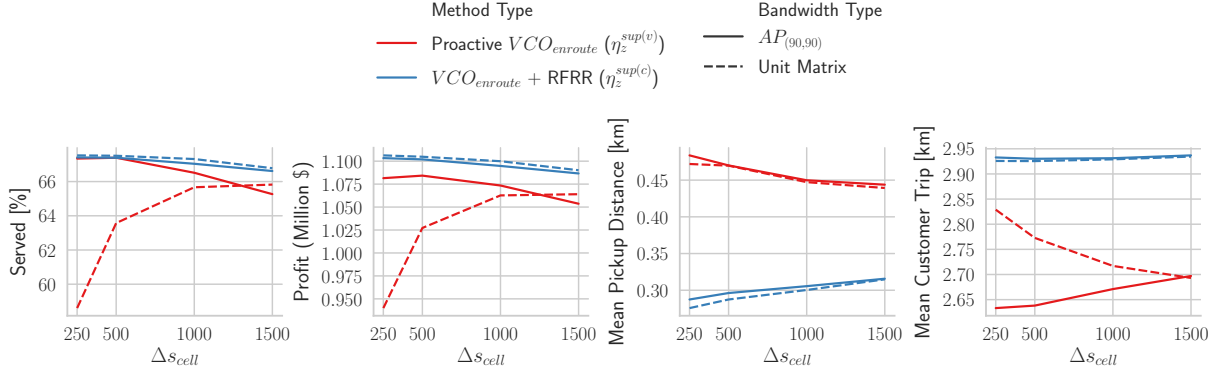
Figure 6.10.: A comparison of AMoD performance for different spatial accuracies of the forecast.

Second, as the value of ΔT_{ph} increases, the AMoD performance decreases. A major reason for this is that the proactive VCO improves the supply-demand imbalance in steps by compromising a small profit in VCO batch, which often results in longer pickup trips as shown in Figure 6.4. As larger ΔT_{ph} forecast longer periods with a smaller probability that the next customer will be immediately found at the destination of enroute vehicle. Thus, longer ΔT_{ph} leads to degraded performance.

Third, the proactive VCO works significantly better when $\eta_z^{sup(v)}$ is used for the supply-side estimate, i.e., when only the customer origins (and not the destinations) are used for calculating the forecast weights. This behavior is quite similar to the behavior of RFRRf method in section 5.6.2; since both methods try to balance the supply-demand ID as much as possible (with the additional constraint of θ in the proactive assignment), the accurate information of the vehicle is important for their performance. The errors in the supply-side estimates in $\eta_z^{sup(c)}$ can significantly degrade the performance. With the usage of $\eta_z^{sup(v)}$, the proactive assignment can distribute the AMoD vehicles more accurately to serve potential customers, improving the overall performance.

Next, the section studies the impact of the spatial accuracy (Δs_{cell}) and the importance of using supply-demand ID in the proactive assignment method. First, as shown in Figure 6.10, a higher spatial accuracy (smaller Δs_{cell}) leads to a significantly higher AMoD performance. With smaller Δs_{cell}, the proactive assignment method can calculate the supply-demand ID with higher accuracy for each VCO batch (Eq. 6.9e) which leads to a better distribution of AMoD vehicles.

To compare the importance of using RFs and ID, the section also considers the AMoD performance when a unit matrix is used in the proactive assignment. Using a unit matrix in Eq. 6.7 makes the formulation independent of neighboring regions; thus, the proactive assignment tries to balance individual regions separately. Figure 6.10 shows that the ID based formulation significantly outperforms the unit matrix-based formulation. The difference is obvious for smaller Δs_{cell}. A smaller Δs_{cell} results in a higher number of regions which the unit matrix-based formulation cannot balance well due to a limited AMoD fleet. By considering the neighboring regions via supply-demand ID, the proactive assignment can position vehicles in a way that leads to better vehicle distribution. However, with the increase in Δs_{cell}, the accuracy of the region-based approximation of ID decreases; since the region-based approximation of ID uses the regional centers for RFs, the bigger regions are not able to correctly represent the reachability of vehicles, leading to a decrease in AMoD performance. Thus, for very high Δs_{cell}, the unit matrix-based formulation performs better than ID based formulation.

Figure 6.10 also shows the performance of RFRR repositioning method (without proactive assignment). Even though with smaller Δs_{cell} the proactive assignment has similar $S_\%$ to RFRR method, most of the served trips are of shorter trip lengths. This still makes the RFRR method perform bet-

ter than a proactive assignment method. However, this also shows that even though the proactive method significantly improves the AMoD performance, an explicit repositioning of vehicles is still essential to improve the performance further.

6.4. Experiments and Sensitivity Analysis for Proactive Assignment Method combined with Repositioning

The previous section showed the performance of the proactive VCO without using it in combination with any of the repositioning methods. When compared against simple VCO (with and without repositioning), it showed significant improvements in AMoD performances, which were already close to the explicit repositioning in chapter 5. However, the previous section showed that there are still chances for further improvements. Therefore, this section studies the AMoD performances when the proactive VCO methods are used together with the explicit repositioning. This section uses the same default simulation configuration as described in Table 6.1. Since chapter 5 already showed that the RFRR method provides the best performance, this section mainly focuses on using the RFRR method together with the proactive VCO. It also compares the performance against the benchmark simulation setup used in the dissertation, i.e., Pavone's repositioning method without the proactive VCO.

6.4.1. Repositioning Frequency

Since the primary purpose of the proactive VCO is to improve the vehicle supply-demand imbalance in small steps via better assignment decisions, it is essential to study if the requirement of the explicit repositioning could be lowered down with proactive VCO. Therefore, Figure 6.11 shows the impact of the varying repositioning period (ΔT_r) on the AMoD performance of proactive VCO.

The first and the most important observation is that a combination of proactive VCO and the repositioning method performance significantly better than using either a proactive VCO (the values with ΔT_r zero in Figure 6.11) or the repositioning separately (the simple $VCO_{enroute}$ with non-zero ΔT_r in Figure 6.11). For example, for ΔT_r of 30 minutes (the default ΔT_r used throughout the dissertation), a combination of proactive $VCO_{enroute}$ and RFRR method serves almost 3% more customers and shows a relative increment of 2.64% compared to the RFRR method without proactive $VCO_{enroute}$. This performance gain is achieved with only 4% repositioning VKT ($D_\%^{repo}$) compared to 9% of without proactive VCO. This shows that the proactive VCO significantly lowers the requirement of explicit repositioning while providing better performance.

Second, the proactive VCO is significantly less affected by higher ΔT_r (lower repositioning frequency) than a simple VCO. For example, in Figure 6.11, when the ΔT_r is increased from 15 minutes to 3 hours, $S_\%$ and the AMoD profit of the simple $VCO_{enroute}$ decreases by 11.2% and 11.4%, respectively. In contrast, for the proactive $VCO_{enroute}$ it decreases by just 0.72% and 0.18%, respectively. This shows that even with a very high repositioning period of 3 hours, the proactive $VCO_{enroute}$ can maintain a very high AMoD performance with $D_\%^{repo}$ of just 1.2%.

Third, since the simple VCO only uses the repositioning method for improving the supply-demand imbalance, a lower ΔT_r generally provides the best AMoD performance — 15 minutes in Figure 6.11. In contrast, a lower value of ΔT_r does not necessarily provide the best performance for the proactive VCO; a very small ΔT_r keeps the vehicles unnecessarily busy with repositioning while a very large ΔT_r does not use the repositioning to its full potential. A moderate ΔT_r provides the best performance for proactive VCO, which in Figure 6.11 corresponds to ΔT_r of 1 hour. However, the proactive VCO is significantly less susceptible to ΔT_r, and other values of ΔT_r still perform very well.
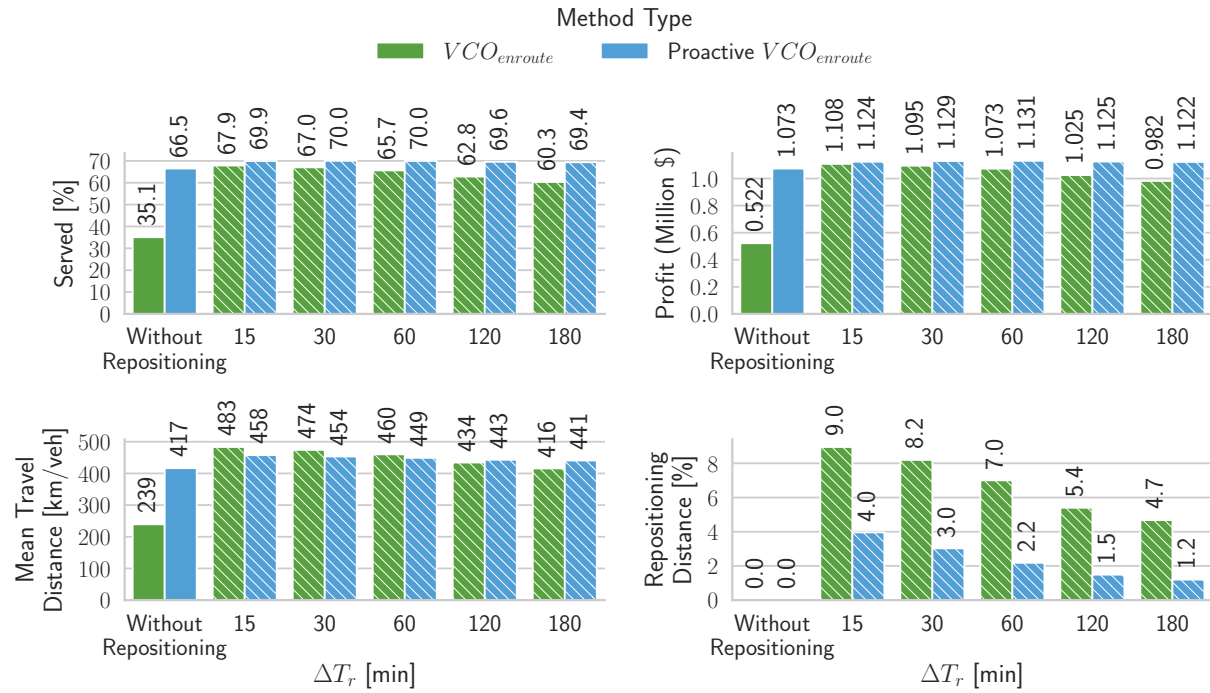
Figure 6.11.: The impact of repositioning frequency (ΔT_r) on the performance of proactive assignment combined with RFRR method.

6.4.2. Maximum Waiting Time

Since the maximum waiting time allowed for picking up a customer (ΔT_{max}) plays a major role in the VCO, the section studies its impact on the proactive VCO. As shown in Figure 6.12, using a tighter time constraint ΔT_{max} can significantly affect the overall performance.

It is observed that for a very small ΔT_{max} the proactive VCO (without repositioning) shows smaller improvement over simple VCO: for ΔT_{max} of 2 minutes, it has a difference of 15.4% in $S_\%$ and a 74.3% increment in AMoD profit compared to a difference of 29.4% in $S_\%$ and an increment of 115.7% in profit for ΔT_{max} of 4 minutes. With smaller ΔT_{max}, the AMoD vehicles have a smaller time window to serve the next potential customers, which can lead to comparatively smaller service quality improvement using proactive VCO. Additionally, a Δs_{cell} of 1 km is used for the experiments in Figure 6.12. With realistic travel times and ΔT_{max} of 2 minutes, the AMoD vehicles cannot reach all the nodes in a region. Thus, the region-based supply-demand ID estimate will have higher errors with such a combination of parameters. With larger ΔT_{max}, the proactive method can calculate the supply-demand ID estimate more accurately and better distribute the AMoD vehicles. When the proactive VCO is combined with the repositioning method for ΔT_{max} of 2 minutes, the AMoD performance is significantly improved. However, the performance is similar to or worse than the combination of simple VCO and the RFRR method for small ΔT_{max}.

In contrast to the above, for a moderate value of ΔT_{max}, combining the proactive VCO and the RFRR repositioning method provides the best performance. The performance could be improved even further by tuning the other parameters of the combination, such as ΔT_r, Δs_{cell}, and θ. It also shows a significant improvement over the benchmark (Pavone's method without proactive VCO). Compared to a pure explicit repositioning (both Pavone's and RFRR method), the combination of proactive VCO and RFRR also reduces the repositioning VKT for all values of ΔT_{max}.
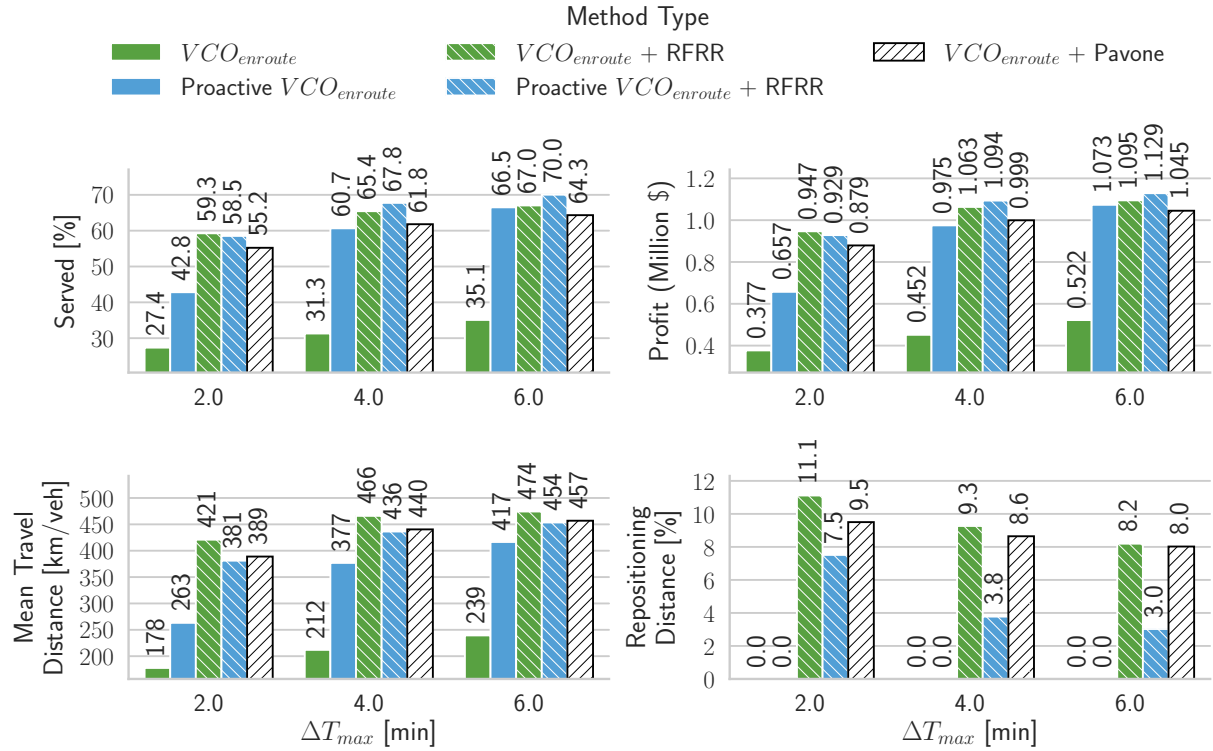
Figure 6.12.: The impact of maximum waiting time allowed (ΔT_{max}) on the combination of proactive assignment and repositioning methods.

6.4.3. Fleet Sizing

This section compares the performances of proactive VCO when combined with explicit repositioning methods and varying fleet sizes. Figure 6.13 compares the AMoD performance of proactive VCO combined with RFRR for mid- and large-sized fleets. Figure 6.14 compares the different proactive VCO types when combined with RFRR for the same fleet sizes. Finally, Figure 6.15 shows the AMoD performance of proactive VCO when combined with min-distance and Pavone's method and a wider range of fleet sizes. The following summarizes the most important observations:

- Overall, the explicit repositioning further improves the performance of proactive VCO for all fleet sizes. Additionally, the improvement is generally higher for small and mid-size fleets. For larger fleet with a very high $S_\%$, the benefits of combining proactive VCO and repositioning decreases as observed for the fleet of above 4500 vehicles in Figure 6.13; rather, the proactive VCO performs worse than simple VCO for a fleet of 6000 and 7500 vehicles. However, it should also be noted that this section uses specific values for parameters; the AMoD performance of a larger fleet can be improved by further tuning θ and ΔT_r. Since large fleets can serve most of the customers when combined with repositioning, the proactive VCO should compromise less on the batch profit (smaller value of θ) and increase the repositioning period ΔT_r (this section uses 30 minutes for ΔT_r).

- The proactive VCO reduces the repositioning distances for all fleet sizes. However, the proactive VCO has slightly higher W_{mean} than simple VCO when combined with the repositioning methods. This is due to the general characteristic of the proactive VCO that increases the
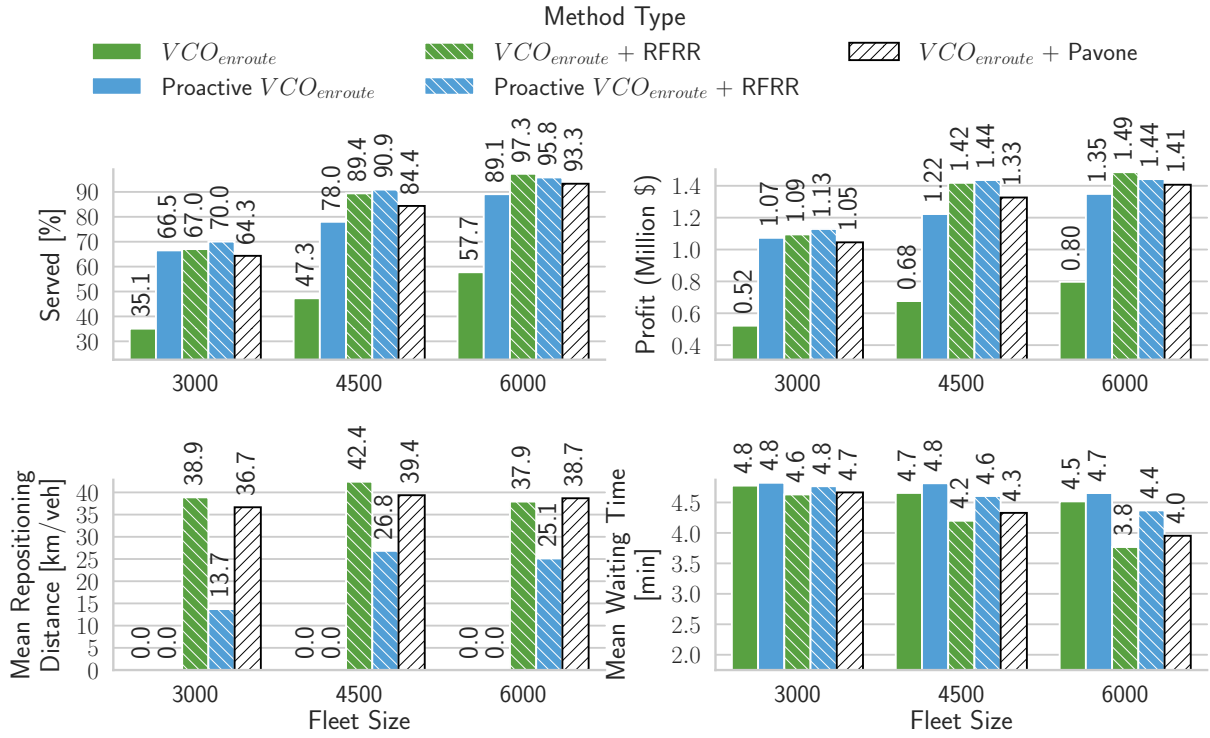
Figure 6.13.: The AMoD performance of varying fleet size using proactive assignment combined with repositioning. The fleet sizes of 4500 and 6000 use θ of 5% while others 20%.

pickup distances and waiting times due to compromising the batch profit, as discussed in section 6.3.1.

- As shown in Figure 6.14 and Figure 6.15, compared to proactive VCO$_{enroute}$, the proactive VCO$_{idle}$ shows greater performance improvement over simple VCO$_{idle}$ when combined with the repositioning method. On average, the combination of proactive VCO$_{idle}$ and RFRR method, serves 3.1% more customers and increases the AMoD profit by 6.24% relative to simple VCO$_{idle}$ combined with RFRR method. In comparison, the combination of proactive VCO$_{enroute}$ and RFRR method serves 0.97% more customers and increases the profit by 0.94% relative to simple VCO$_{enroute}$ combined with RFRR method. The rest of the observations are the same as above for the proactive VCO$_{idle}$ compared to simple VCO$_{idle}$.

- In regards to the performance of other repositioning methods with the proactive VCO, the min-distance provides slightly better performance than RFRR for small fleets (1500 vehicles) as shown in Figure 6.15. However, for mid- and large-sized fleets, the RFRR method provides the best performance.

- Compared to the benchmark method used in the dissertation, i.e. simple VCO with Pavone's repositioning, the combination of proactive VCO and RFRR repositioning significantly improves the AMoD performance. For the VCO$_{idle}$, on average for all fleet sizes, it serves 4.65% more customers (a relative increase of 12.6%) with a 9.95% relative increase in profit. The improvement is higher for smaller fleets; for a fleet of 1500 vehicles, there is a relative increment of 29% in AMoD profit while 8.1% and -1.1% for a fleet of 4500 and 6000 vehicles, respectively. Similarly, on average, the combination of proactive VCO$_{enroute}$ and RFRR serves 3.95% more
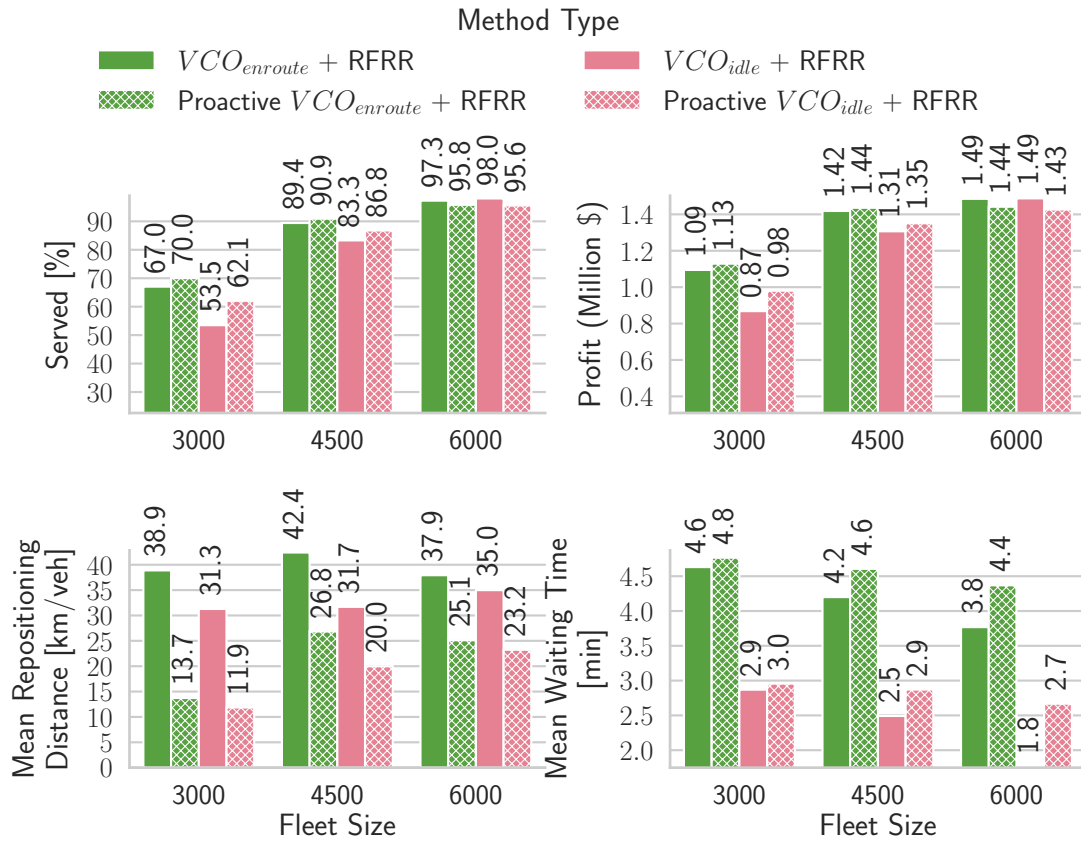
Figure 6.14.: The comparison proactive assignment combined with repositioning for different VCO types. The fleet sizes of 4500 and 6000 use θ of 5% while others 20%.

customers (a relative increase of 6.6%) with a relative profit increment of 5.9%. The fleet of 1500 vehicles has a relative profit increment of 11.24% which reduces to 8.2% and 2.48% for a fleet of 4500 and 6000 vehicles, respectively.

- Even without repositioning, the proactive VCO outperforms the combination of explicit repositioning with non-proactive VCO for small and mid-sized fleets (up to a fleet of 3000 vehicles), especially compared to Pavone's method; for example, the proactive $VCO_{enroute}$ with 3000 vehicles produces a profit of $1.07 million compared $1.04 million using Pavone's method and non-proactive $VCO_{enroute}$. This shows the effectiveness of using the proactive assignment approach.

In summary, the results show the potential benefits of using the proactive VCO method. A significant performance improvement is observed even if the proactive VCO is used without explicitly repositioning idle vehicles. However, combining the proactive VCO with the repositioning method shows further AMoD performance improvement. It even outperforms the AMoD performances observed in chapter 5, which could be further improved by tuning the parameters of the methods used. The improvements are also found to be consistent for multiple fleet sizes.

(a) VCO$_{idle}$

(b) VCO$_{enroute}$
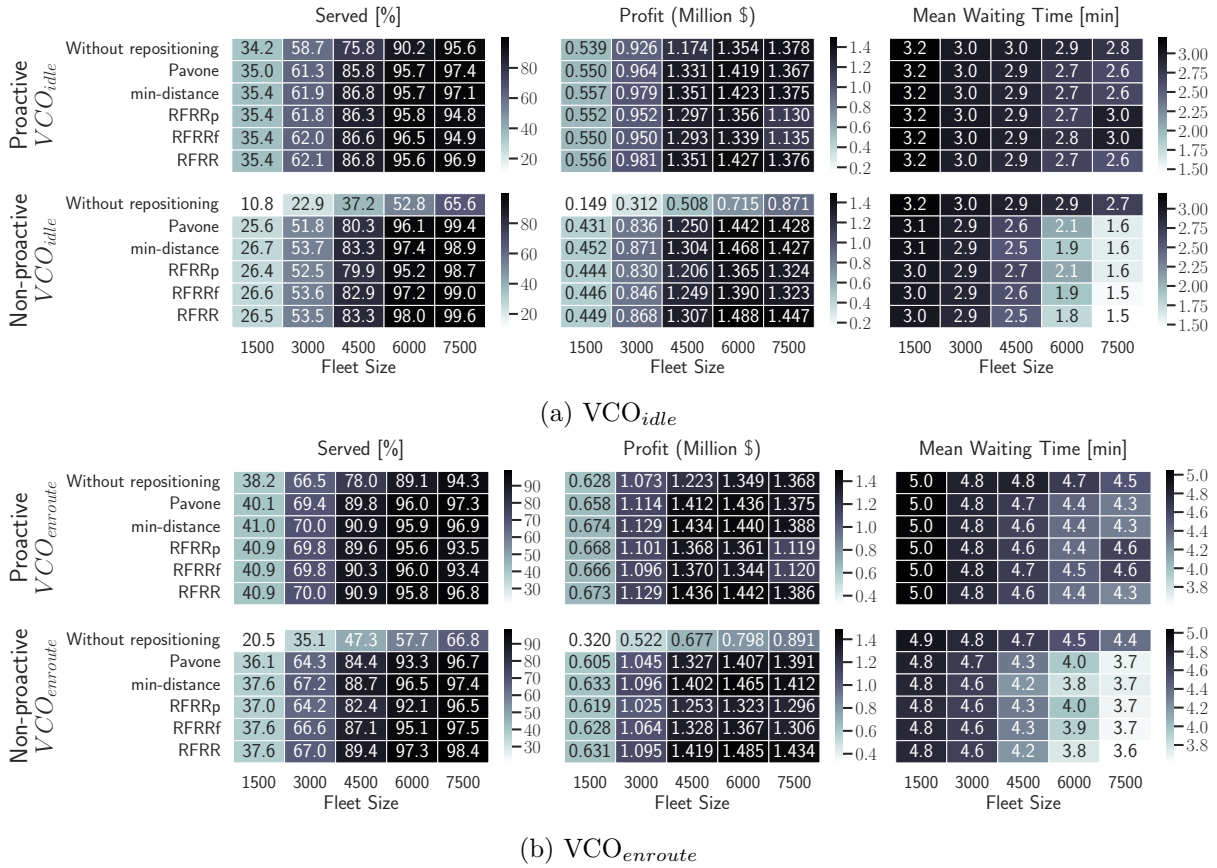
Figure 6.15.: The AMoD performance of combination of proactive VCO with multiple repositioning methods. The fleet sizes of 4500 and 6000 use θ of 5% while others 20%. The RFRRp and RFRRf use $\omega_z^{sup(v)}$ for vehicle supply estimate and γ of 0.7.

6.5. Conclusion

This chapter presented a proactive assignment method for the AMoD services. The method compliments VCO and can be used with any VCO method. Primarily, it compromises a percentage of the original objective function of VCO to achieve long-term supply-demand balance in small steps. The chapter formulated for both VCO types used in the dissertation. Even though the proactive assignment method can be used without ID, the results indicated that using region-based estimation to ID in the proactive assignment method provides higher performance, especially when the operation area is divided into smaller regions. The results show that the proactive VCO significantly improves the AMoD performance without the need for explicit repositioning of idle vehicles — in some cases, it even outperforms repositioning methods with non-proactive VCO. The performance of proactive VCO increases further when used together with repositioning methods. This additionally allows a higher repositioning period as the proactive VCO already reduces the need to reposition idle vehicles explicitly.

This chapter concludes the main contributions of the dissertation. The next chapter provides the overall summary and conclusion of the dissertation.

Chapter 7.

Conclusion

This chapter summarizes the overall findings of the dissertation by answering the research questions raised in section 1.3. It also discusses the limitations of the dissertation that must be observed when using the results. Finally, it describes the potential areas for future extensions of the methods developed in the dissertation.

7.1. Summary of the Research

To improve the efficiency and performance of AMoD services, the dissertation is divided into three main contributions: (1) the development of a novel spatiotemporal relation to describe AMoD supply-demand imbalance (chapter 4), (2) the development of novel repositioning algorithms based on the newly developed spatiotemporal relation (chapter 5) and (3) the development of a novel VCO technique that uses the spatiotemporal relation for proactively minimizing the supply-demand imbalance (chapter 6). Each chapter answered the research questions of section 1.3. The following discusses their main findings and how they answered the related questions.

7.1.1. Novel Spatiotemporal Relationship

The first significant research gap focused on in the dissertation is the lack of a consistent spatiotemporal relationship between the AMoD supply and demand. Often, the literature divides the operation area into a disjoint set of regions to measure the supply-demand imbalance in each region. This presupposes that the individual regions are independent such that an AMoD vehicle from one region only serves the customers in that region. Some researchers tried to solve this problem using hierarchical regions. Nevertheless, there is still a lack of a consistent relationship that could describe a spatiotemporal relationship between supply and demand. Therefore, chapter 4 focused on the following two questions:

- **RQ 1.1**: How can a vehicle distribution model be defined that is independent or at least loosely bound to regional shapes?

- **RQ 1.2**: Does the above distribution model has a strong and consistent relationship between states of individual regions and the overall AMoD system? How can it be measured?

To answer RQ 1.1, chapter 4 took inspiration from KDE to build a novel spatiotemporal metric for the AMoD services. KDE is a commonly used non-parametric statistical tool that uses kernel functions at each data point to estimate the underlying pdf. To make the relationship independent of regions, chapter 4 modified the kernel functions to reachability functions (RFs). An RF uses the maximum allowed customer waiting time to calculate the reachable distance (bandwidth) of an RF, in contrast to a KDE where all kernels use the same bandwidths. This helped to define supply-demand

imbalance in the form of a imbalance density (ID) function. An ID directly uses the vehicle and customer pickup locations independent of operational regions.

To answer RQ 1.2, chapter 4 used agent-based simulation in the Manhattan area using NYC data. It first simulated the AMoD scenario with different VCO methods and calculated how well a disjoint set of independent regions and ID describe the VCO decisions. The simulation data of each VCO batch was used separately to observe which relationship better describes the AMoD operation. The chapter followed the procedure that simulation data is only passively used in the analysis, guaranteeing any relationship observed between AMoD supply and demand is not due to the manipulation of the VCO decisions by the underlying assumption being observed. The following summarizes the main findings:

- Regarding the assumption of independent regions, the simulation results showed that many vehicles serve customers from the neighboring regions, especially when the regions are small and contiguous. This confirmed the primary hypothesis of the research, i.e., the assumption of independent regions does not accurately describe the AMoD operation.

- The ID relation accurately describes the VCO decisions in terms of customer waiting times and probability of being served by an AMoD vehicle. However, the strength of the relationship in describing the service quality at each geographical location also depends on the VCO method used. The VCO_{idle} method that only uses the idle vehicle for the assignment showed a strong relationship between customer waiting times and the probability of being served by an AMoD vehicle. In contrast, the $VCO_{enroute}$ only showed a strong relationship with the probability of being served.

- The ID also provides a novel way of plotting the offered AMoD service quality in the form of a heat map. The heat map helps identify service quality provided in different parts of the operation area; the parts with negative ID marks the areas with a deficiency of vehicles, while the parts with positive ID marks the areas with surplus vehicles. The magnitude of ID also indicates the level of deficiency or surplus. Thus, the ID serves as an essential spatiotemporal metric for the AMoD service quality. The chapter asserts that any method that actively minimizes the supply-demand ID will improve the AMoD service quality.

7.1.2. Repositioning Method

The assumption of independent regions is mostly used with other AMoD algorithms to improve service quality. Among these, the most famous are the algorithms that periodically reposition idle vehicles to higher-demand regions. This resolves the supply-demand imbalance accumulated over time. The dissertation suggested that the assumption of independent regions limits tapping the full potential of repositioning algorithms. Thus, chapter 5 aimed to answer the following question:

- **RQ 2.1**: How can the newly developed vehicle distribution model be incorporated into the repositioning problem?

To answer the above question, chapter 5 developed novel repositioning methods that incorporate the ID in its formulation. For this purpose, it first developed a general repositioning formulation that directly used ID based on raw coordinates of vehicle supply and demand. However, such a repositioning formulation leads to impractical computational complexity. Additionally, vehicle and customer forecasts are usually unavailable on such a granularity. Usually, the forecasts are available on a regional level. Thus, instead of using the ID based on raw coordinates, chapter 5 used its

region-based estimation. Accordingly, three reachability function based repositioning (RFR) methods were developed depending on the regional restrictions on the origins and destinations of repositioning vehicles: RFRR, RFRRp and RFRRf. The RFRR only allowed repositioning from surplus to deficiency regions, RFRRp additionally allowed repositioning from surplus to all regions, and RFRRf did not have any regional restriction. Thus, RFRRf optimized the ID as much as possible, followed by RFRRp and RFRR, respectively.

Chapter 5 tested the performances of RFR methods against Pavone's method [PAVONE et al., 2012]. Although Pavone's method is old, the method is often used in the AMoD literature. Additionally, it provided the best performance among other methods tested in the dissertation[1]. Pavone's method uses the assumption of independent regions and tries to distribute the excess vehicles equally among regions. Additionally, chapter 5 also developed a separate independent regions based formulation called the min-distance method that sent just enough vehicles to each region that satisfies the regional demand. Although the RFR methods also used region-based formulation, the key difference between RFR and independent regions based methods is that the RFR considered the impact of neighboring regions in repositioning decisions. The following summarizes the key findings:

- The consideration of ID density in repositioning decisions significantly improves AMoD performance.

- Despite the performance improvement by using ID, the unrestricted minimization of estimated ID can lead to a deteriorated performance. Thus, RFRRp and RFRRf are observed to perform worse than RFRR method. The estimation of long-term vehicle supply plays a significant role in this behavior. Since the RFRRp and RFRRf have a higher emphasis on optimizing the ID, the consequences of wrong repositioning decisions are more severe in comparison to RFRR method, which mitigates the adverse effects by using regional restrictions. This is especially problematic during the night hours when there is a significant increase in idle vehicles and the RFRRp and RFRRf methods lead to a high repositioning VKT to balance the ID. This also makes the RFRRp and RFRRf perform better when the customer destination points are not used for estimating the long-term supply of vehicles.

- In a ID-based analysis of VCO batches (similar to the one performed in chapter 4), it is found that even though the min-distance method does not directly optimize the long-term ID, it still leads many customers to be on the positive side of ID metric. This makes the min-distance method provide the second best AMoD performance after RFRR method. This also shows that even the methods that do not necessarily consider ID in their formulation can still lead to a good distribution of ID values measured for individual VCO batches. The main reason for this is that unlike the ID measured for individual VCO batches, the repositioning formulation deal with long-term forecast values and estimation of future ID, which is not necessarily equivalent to the actual ID values of each VCO batch. This behavior is further affected by the repositioning of vehicles taking some time to reach their destinations.

7.1.3. Proactive Vehicle Assignment Method

While chapter 5 mainly focused on improving the AMoD services by efficient repositioning of idle vehicles, chapter 6 focused on proactively mitigating the emergence of supply-demand imbalance. This means that the vehicles are assigned to customers in a way that improves the long-term supply-demand imbalance. This would also reduce the need for frequent repositioning of idle vehicles. Thus, the chapter focused on two main questions:

---

[1]Appendix B presents the formulations used for the benchmark methods.

- **RQ 3.1**: How to incorporate the statistical information inside a VCO?

- **RQ 3.2**: How to utilize the vehicle distribution model developed in RQ 1.1 into the proactive VCO?

To answer RQ 3.1, chapter 6 built a proactive assignment method that complements other VCO methods. The fundamental concept of the approach is to compromise a percentage of the VCO objective (i.e., monetary profit in the dissertation) to get an assignment of vehicles that mitigates the supply-demand imbalance. This raised a secondary question of how to measure the impact of each VCO decision on the supply-demand imbalance. To answer this question, chapter 6 used the VCO decision variables in supply-demand imbalance measurement. Then, to address RQ 3.2, it incorporated the region-based ID in the newly developed formulation. The following summarizes the main findings:

- The proactive VCO schemes significantly improves the AMoD performance. Even a small compromise of the batch profit ranging from 5 to 20% —depending on the fleet size— can improve the AMoD profit by 60 to 100%.

- The proactive VCO was tested with both types of VCO methods used in the dissertation, i.e. VCO_{idle} and $VCO_{enroute}$. It improves the AMoD performance in both cases, indicating the complementary nature of the method.

- To measure the benefits of using ID in the formulation, the proactive VCO was also tested using a formulation based on the assumption of independent regions. The incorporation of ID into the formulation significantly improves the AMoD performance, especially when the region sizes are small. This is due to a higher accuracy of ID estimation for smaller regions.

- The proactive VCO provides significantly better performance when the vehicle supply forecast does not utilize the future customers' destinations. This behavior is similar to the RFRRf method. Since both approaches try to distribute the vehicles in a way that reduces the ID as much as possible, any error in the vehicle-supply estimate deteriorates the performance.

- The proactive VCO reduces the supply-demand imbalance while serving customer requests. Some idle vehicles would still require explicit repositioning for even higher performance. Thus, the dissertation found that the combination of proactive VCO and the repositioning methods (especially the RFRR method) provides the best AMoD performance. It was also found that with proactive VCO, even a very low-frequency repositioning frequency does not significantly affect performance. This shows that with proactive VCO, the requirement of explicit repositioning is reduced considerably.

7.1.4. Overall AMoD improvement

As described in section 1.3, the main question researched in the dissertation is the following:

- MRQ: How to improve the overall AMoD performance using novel FC methods?

By developing the novel FC methods based on spatiotemporal relation, the dissertation answered the MRQ. Combining the proactive VCO and the RFRR repositioning method significantly improves over the benchmark method, i.e., Pavone's repositioning method without proactive assignment. Table 7.1 summarizes the overall performance gain achieved using the techniques developed in the

Compared to	Method	3000 vehicles		Average for all fleet sizes	
		Customers Served	Porfit	Customers Served	Porfit
Without repositioning and proactive assignment	Proactive Assignment (Without Repositioning)	89.6%	105.6%	67.3%	81%
	RFRR Repositioning	91.1%	109.8%	75.9%	92.7%
	Proactive Assignment and RFRR repositioning	99.6%	116.3%	80.5%	95%
Pavone's repositioning method without proactive assignment	Proactive Assignment (Without Repositioning)	3.4%	2.7%	-1.0%	-1.4%
	RFRR Repositioning	4.2%	4.8%	4.1%	4.9%
	Proactive Assignment and RFRR repositioning	8.9%	8%	6.6%	5.9%

Table 7.1.: Summary of the overall AMoD performance gain. The values show percentage improvement over the simulation configuration given in the left column using $\text{VCO}_{enroute}$. The improvement for the default configuration used (i.e., a fleet of 3000 vehicles) is shown separately.

dissertation. Compared to a simple AMoD scenario without repositioning, on average, for all fleet sizes, the proactive assignment serves 67.3% more customers with an 81% increase in the AMoD profit. This improvement is crucial for the AMoD services, especially when no explicit repositioning of idle vehicles is required to achieve it. For mid-sized fleets (for example, 3000 vehicles), the performance of the proactive assignment is even better than the explicit repositioning of the benchmark method — 3.4% increase in customers served and 2.7% increment in AMoD profit. However, the proactive assignment alone cannot keep up with the benchmark method for larger fleets and performs worse on average for all fleet sizes than the benchmark method.

The introduced repositioning method RFRR also performs significantly better than the benchmark method, serving 4.2% more customers and producing 4.8% more profit for a fleet of 3000 vehicles. On average, it serves 4.1% more customers for all fleets with a profit gain of 4.9%. The most important outcome of the dissertation is produced when the proactive assignment is combined with the repositioning method RFRR. It doubles the performance gain of RFRR over the benchmark method for mid-sized fleets, serving 8.9% more customers and producing 8% more profit. On average, for all fleets, this combination serves 6.6% more customers with an improvement in AMoD profit of 5.9%. Such performance improvement may appear small at first sight, but it must be noted that these improvements are just for two days of simulation. When projected for a whole year, this produces a significant monetary gain. For example, for a fleet of 3000 vehicles, this means a daily gain of 42 thousand dollars and a yearly gain of almost 15.3 million dollars.

This concludes the summary of the main findings of the dissertation. The following section discusses the main limitations and potential of future research.

7.2. Limitations and Future Research

While the dissertation attempted to use a realistic AMoD simulation for testing the methods, some underlying assumptions may limit their application and must be considered when interpreting the results. The following provides the list of the most important limitations and the main areas of

improvement for future research:

- The methods were developed and tested for AMoD service without pooling customers; thus, pooling of customers can significantly affect the performance and may require modifications of the methods. Therefore, future research should focus on extending the RFs based spatiotemporal metric to the pooling of customers in AMoD services. Accordingly, future research should develop the RFR and the proactive assignment methods to pool customers.

- The dissertation used a simplistic customer model where each AMoD offer is accepted by the customer. In reality, the decision to accept or reject the AMoD offer is a complicated process depending on multiple factors. Some methods introduced, especially proactive VCO, may perform differently when tested with a more realistic customer model. Thus, in the future, the methods should be tested with more advanced customer models. This is especially important for the proactive assignment approach, where some customers may have a higher waiting time or do not even get an AMoD offer based on their destination location. It is essential to include these factors in the customer's decision to accept or reject the offers or even in deciding whether to use the AMoD service in the future. On the AMoD side, the performance can be affected by late-arrival, no-shows, or cancellation of the requested ride, which should be investigated in the future.

- Even though the proactive assignment method leads to significant improvement in AMoD performance, the strategy does not consider the equal treatment of all customer requests regardless of their destination locations. This may hinder its use in an actual AMoD service. Therefore, future research can focus on a dynamic pricing strategy that compensates the customers via monetary gains or coupons. In this regard, interesting research would be to combine proactive VCO with block-chain based currency like Mobility Coin[BLUM et al., 2022], which would allow the customers to utilize the compensation points in other modes of transport.

- The simulation framework used in the dissertation only considered a single AMoD service. In reality, similar to current MoD services, the future AMoD services will also have AMoD competitors operating in the same area. This limits the application of some of the FC methods used in the dissertation. Thus, in the future, the methods must be tested in a simulation environment where the customers can select among multiple AMoD operators.

- The dissertation used AMoD simulation in Manhattan area using NYC data. Compared to other cities, the Manhattan area is small, with a very high concentration of trips. Usually, the performance of an FC method depends on the overall vehicle supply and demand ratio in the data used. The FC methods introduced may perform differently when tested on other data sources and cities. Thus, future research can apply the introduced methods to other data sets and cities.

- The dissertation used two specific VCO methods to assign vehicles to customers. However, the literature contains multiple VCO methods, including the ones based on the reoptimization of already assigned customers. These VCO methods can also significantly impact the performance and can be investigated in the future.

- For the calculation of RF bandwidths, the dissertation developed and used AP algorithm. However, a different approach for their calculation may significantly affect performance. Additionally, the travel times only changed every 30 minutes in the simulation environment, allowing

the preprocessing of the RF bandwidths for each period. Future research can focus on developing a method that can calculate RF bandwidth in real-time using more accurate network data. This would allow the usage of the RF based methods in real AMoD services.

- The dissertation used NYC data to derive realistic travel times that change every 30 minutes. The effects of congestion were implicitly included by scaling each network edge. Nevertheless, the simulation did not use a microscopic traffic simulation with realistic congestion effects, which may be focused on in future research.

- The simulations also did not consider additional maintenance tasks like cleaning and refueling vehicles, which can also affect performance. These can also be investigated in the future.

In summary, the AMoD services are still in the pilot phase, and the above assumption had to be made to focus on the research questions. Overall, the methods introduced improved the AMoD performance significantly, which is expected to perform even better with the above suggestions for future research.

List of Figures

List of Tables

Mathematical Notations

The following table summarizes the mathematical notations used throughout the dissertation. For clarity, any symbol used only in one section is not included in the list; thus, those symbols are only explained in the respective section. Additionally, some notations, like the decision variables for an optimization problem, are reused in multiple sections and hence, not included in the following list. If a notation represents a parameter requiring a numerical value, the following table lists its default value used in the dissertation. The table also includes the section where the notation is mainly defined. For a detailed description of the notation, refer to the related section.

Symbol	Description	Default Value	Section
c^D	The distance-based cost of an AMoD vehicle.	$0.25 per km	3.1.2
c^F	The fixed daily cost of an AMoD vehicle.	$25 per vehicle	3.1.2
c^{VOT}	The monetary cost for the value of time of customers (not used in the dissertation).		3.1.3
d_r	The drop-off location of a customer $r \in R$.		3.1
d_r^{pd}	The distance from pickup location p_r to drop-off location d_r of a customer $r \in R$		3.1.2
D_{total}^{empty}	The total empty distance driven by AMoD fleet in a single simulation run.		3.1.4
$D_{\%}^{empty}$	The percentage of empty distance out of total distance driven by the AMoD fleet.		3.1.4
$D_{\%}^{pick}$	The percentage of the total distance driven by AMoD fleet for picking up customers.		3.1.4
D_{mean}^{pick}	The mean pickup distance driven by AMoD fleet for picking up customers.		3.1.4
D_{total}^{pick}	The total distance driven by AMoD fleet for picking up customers.		3.1.4
$D_{\%}^{repo}$	The percentage of the total distance driven by AMoD fleet for repositioning of idle vehicles to demand-intensive regions.		3.1.4
E	The edges of the bipartite graph G connecting vehicles and customer requests. An edge only exists if the vehicle can pick up the customer within the time-window $[t_r, t_r + \Delta T_{max}]$.		3.1.3
E_{net}	The edges of the graph G_{net}.		4.2.4
f^D	The distance-based variable fare charged from the customer.	$0.5 per km	3.1.2
G	The bipartite graph of feasible assignments between vehicles and customer requests used for vehicle control optimization problem.		3.1.3

Symbol	Description	Default Value	Section
G_{net}	The graph of the city network used.		4.2.4
h	The bandwidth of the kernel function used in KDE. The dissertation uses it to measure reachable distances from a point in reachability functions.		4.2.1, 4.2.2
$K(x)$	A smooth kernel function usually used in KDE. The dissertation uses it as a reachability function.	2D triangular function	4.2.1, 4.2.2
$k(x, x_i, h)$	A kernel function $K(x)$ with an offset. The notation means that a kernel function $K(x)$ is centered at the point x_i having a bandwidth h.		4.2.1
$K_{\geq 0}$	Integral of $\bar{p}_Z(x)$ over the domain with positive values of $\bar{p}_Z(x)$.		5.5.2
$K_{\leq 0}$	Integral of $\bar{p}_Z(x)$ over the domain with negative values of $\bar{p}_Z(x)$.		5.5.2
N_{net}	The nodes of the city network graph G_{net}.		4.2.4
$\bar{p}(x)$	The reachability density function that inter-connects individual geographical points using reachability functions.		4.2.2
$\bar{p}_Z(x)$	Region-based approximation to reachability density function $\bar{p}(x)$ using regions Z. The centroids of the regions in Z are used as data points.		5.4.2
p_r	The pickup location of a customer $r \in R$.		3.1
R	The set of all customer requests.		3.1
$S_\%$	The percentage of the customers served.		3.1.4
t_r	The time at which a single customer $r \in R$ requests a ride.		3.1
$U_\%$	The fleet utilization is defined as the percentage of time the AMoD fleet was busy. It can also be defined for specific tasks like customer pickups or repositioning, represented by the specific task mentioned after $U_\%$.		3.1.4
V	The fleet of AMoD vehicles.	3000 vehicles	3.1
W_{mean}	The mean waiting time of picking up customers.		3.1.4
W_Z	The set of all forecast weights ω_z for the regions Z.		5.2.2
x	A single data point used in the kernel or reachability function. In the dissertation, it represents a 2D coordinate on the map.		4.2.1, 4.2.2
Z	The set of all the regions in the area of operation. In the dissertation, this refers to the set of regions obtained using a regular grid.		3.3.2, 5.2.2
γ	The relative importance of the supply-demand imbalance objective to the travel distance objective for reachability functions based repositioning (RFR) methods.		5.4.4
ΔK_{abs}	The change in the absolute values of $K_{\geq 0}$ and $K_{\leq 0}$ caused by repositioning.		5.5.2

Symbol	Description	Default Value	Section
Δs_{cell}	The size of a single cell in the regular grid. It specifies the regular grid used for generating the AMoD operational regions. Δs_{cell} is also used spatio-temporal scaling of the travel times of the network.	1 km	3.2.3
ΔT_{batch}	The batching period used for accumulating customer requests before solving a vehicle control optimization problem.	30 seconds	3.1.3
ΔT_h	The time horizon of AMoD supply and demand forecast used for repositioning of idle vehicles.	30 minutes	5.2.2
ΔT_{max}	The maximum waiting time starting from t_r allowed to pick up a single customer	6 minutes	3.1
ΔT_{ph}	The time horizon of AMoD supply and demand forecast used for proactive vehicle control optimization method.	15 minutes	6.2.2
ΔT_r	The time period used for repositioning of idle vehicles.	30 minutes	5.2.2
ΔT_{scale}	The period used for grouping the historical trips for scaling the travel times of the network.	30 minutes	3.3
ζ	The base fare charged per served customer.	\$2.5 per customer	3.1.2
$\boldsymbol{\eta}$	The set of all forecast weights η_z for the regions Z used for proactive vehicle control optimization method.		6.2.2
η_z	The forecast weight of zone $z \in Z$ used for proactive vehicle control optimization method, given as $\eta_z = \eta_z^{sup} - \eta_z^{dem}$.		6.2.2
η_z^{dem}	The weight for customer demand forecast in zone $z \in Z$ used for proactive vehicle control optimization method.		6.2.2
η_z^{sup}	The weight for vehicle supply forecast in zone $z \in Z$ used for proactive vehicle control optimization method.		6.2.2
$\eta_z^{sup(c)}$	A type of vehicle supply forecast (η_z^{sup}) where customer destination points are also included in the regional weights.		6.2.2
$\eta_z^{sup(v)}$	A type of vehicle supply forecast (η_z^{sup}) where only the current information of vehicle routes is used for regional weights.		6.2.2
$\rho_{imb}(x)$	An imbalance density function for linking the supply-demand imbalance using raw coordinates and reachability functions.		4.2.3
ω_z	The forecast weight of zone $z \in Z$ used for repositioning, given as $\omega_z = \omega_z^{sup} - \omega_z^{dem}$.		5.2.2
ω_z^{dem}	The forecast weight for customer demand in zone $z \in Z$ used for repositioning.		5.2.2

Symbol	Description	Default Value	Section
ω_z^{sup}	The forecast weight for vehicle supply in zone $z \in Z$ used for repositioning.		5.2.2
$\omega_z^{sup(c)}$	A type of vehicle supply forecast (ω_z^{sup}) where customer destination points are also included in the regional weights.		5.2.2
$\omega_z^{sup(v)}$	A type of vehicle supply forecast (ω_z^{sup}) where only the current information of vehicle routes is used for regional weights.		5.2.2

List of Terms and Abbreviations

VCO$_{enroute}$ VCO using idle and en-route vehicles 43, 44, 46, 73–76, 79–81, 83, 86, 87, 110, 112–117, 119, 123–132, 140, 142–146, 148, 151–153, 156, 158, 159, 164–167, 203

VCO$_{idle}$ VCO using idle vehicles only 43, 73–83, 111, 124, 125, 128–131, 140, 146, 151, 153, 156, 158, 164–166

AMoD   autonomous mobility on-demand iii, iv, 2–8, 11, 12, 15, 17–42, 44–48, 50, 54, 55, 59, 62–69, 72, 73, 75–77, 79, 81–83, 85–89, 92, 93, 101, 102, 107, 109–113, 116, 118–129, 131–133, 135–138, 143–153, 155–161, 163–167, 169–171, 195, 203, 204

AP   adaptive percentile 70–72, 126, 160

ASM   absolute scaling method 57, 59–62, 73, 75–78, 80, 82, 86, 87, 110, 142, 164

AV   autonomous vehicle iii, 2–5, 7, 11, 12, 20–25, 27, 28, 36, 38, 39, 42, 46, 73, 82, 87, 110, 142, 164

B&B   branch-and-bound 16, 17

B&C   branch-and-cut 17

CI   charging infrastructure 24, 25, 38, 39

CS   car-sharing 21, 23, 25, 27, 29, 37

CTS   Cybernetics Transportation System 21

CVRP   Capacitated Vehicle Routing problem 33

DARP   Dial-a-Ride problem 23, 33–36

DP   dynamic programming 39

eAMoD   electric AMOD 24, 25, 38, 39, 42

EV   electric vehicle 24

FC   fleet controller iii, 3–6, 8, 24, 28, 32–36, 38, 42, 46–48, 51, 53, 62, 65, 75, 76, 87, 93, 125, 132, 135, 137, 158, 160

FCFS   first come first serve 35, 38

FF   free-flow 72

FFT   fast Fourier transform 94

FP   fixed percentile 70–72

GIS   geographical information system 20, 48

GPS   global positioning system 3

ICT	information and communication technologies 21
ID	imbalance density iii, iv, 68, 72, 76, 77, 81–83, 85, 88, 92, 94, 98, 99, 101, 102, 105–107, 111–113, 117–119, 121, 123, 126–128, 132, 133, 135, 137, 144, 147, 149, 153, 156–158, 164, 165
ILP	integer linear programming problem 16, 17
IP	integer programming problem 11, 14–17
IQP	integer quadratic programming problem 16, 17
KDE	kernel density estimation iii, 63–67, 94, 155, 170
KPI	key performance indicator 45, 46, 101, 102, 109, 165
LP	linear programming problem 16, 17
MaaS	mobility as a service 25, 26
MDP	Markov decision process 5, 7
MDPP	minimum drift plus penalty 39
MFD	macroscopic fundamental diagram 27
MFM	mean factor method 54, 55, 59, 60, 62
MIP	mixed integer programming problem 14, 17, 94, 100
ML	machine learning 36
MoD	mobility on-demand iii, 1–5, 20, 23–26, 28, 29, 32, 38, 39, 45, 47, 50, 85, 160, 163, 196
MPC	model predictive control 39
MSP	mobility service provider 1–5, 13, 15, 22, 23, 27, 32, 38, 44, 47, 64, 92, 100, 138, 141
NYC	New York City iv, 2, 7, 28, 36, 41, 46, 49–53, 55, 57, 59, 62, 63, 85, 100, 101, 156, 160, 161, 163
OSM	OpenStreetMaps 7, 20, 48, 49, 54, 55, 57–59, 163
PBD	prioritized balanced density 100, 102–106, 110, 111, 116, 118, 119, 127, 143, 164–166
pdf	probability density function 65–67, 155
PDP	pickup-and-delivery problem 33
POV	privately owned vehicle 1, 2, 4, 24, 25, 39
PT	public transport 1–4, 20, 25, 31
QP	quadratic programming problem 16
RD	reachability density 64, 67, 68, 72, 92, 94, 117, 118, 140
RF	reachability function iii, 65–67, 69–73, 77–79, 83, 85, 88, 93, 94, 97, 101, 107, 110, 120, 123, 124, 126, 127, 139, 142, 147, 155, 160, 161, 164, 165

RFR	reachability function based repositioning iv, 92, 94, 99, 100, 102–108, 110, 112, 119, 120, 124, 126, 132, 135, 138, 139, 157, 160, 164, 165, 170
RFRR	RFR with regions 94, 95, 99, 102, 103, 107–113, 117–120, 122, 123, 125–130, 132, 133, 135, 136, 142–151, 157–159, 164–166
RFRRf	RFR with regions and full flow 96, 98, 99, 102, 103, 107, 108, 110–113, 117–121, 123–128, 131, 132, 144, 145, 147, 153, 157, 158, 165, 166
RFRRp	RFR with regions and positive zones flow 98, 99, 103, 107, 108, 110–112, 117, 118, 123, 125, 128, 130–132, 145, 153, 157, 165, 166
RH	ride-hailing iii, 3–5, 21, 23, 28, 33, 35, 36, 40–42
RL	reinforcement learning 36
RS	ride-sharing iii, 3, 21, 23, 26, 27, 33, 36–40
SAE	Society of Automotive Engineers 20
SAV	shared autonomous vehicle 4
SD-DARP	stochastic and dynamic DARP 34, 35
SDVRP	stochastic and dynamic vehicle routing problem 5, 34–36, 42, 116
SP	stated preference 29
SSM	squared scaling method 56, 57, 59, 60, 62
TNC	transportation network company 1
TSP	travelling salesman problem 19
UTM	Universal Transverse Mercator 57
VCO	vehicle control optimization 5, 7–9, 11, 12, 41–45, 47, 48, 62, 67, 69, 72, 73, 75–81, 87–89, 101, 107, 110–112, 124, 125, 128–130, 133, 135–153, 155–158, 160, 163–166
VE	vehicle emulator 48
VKT	vehicle kilometers travelled 2, 4, 24, 25, 37, 45, 46, 87, 88, 91–94, 96, 98–100, 102, 107–109, 111, 113, 117, 125–128, 135, 136, 144, 148, 149, 157, 201
VRP	vehicle routing problem 19, 20, 32–34, 36, 37
VSP	vehicle saturation point 107, 108

Bibliography

ABKARIAN, HOSEB; HANI S. MAHMASSANI; MICHAEL HYLAND (Aug. 2022). "Modeling the Mixed-Service Fleet Problem of Shared-Use Autonomous Mobility Systems for On-Demand Ridesourcing and Carsharing With Reservations". In: *Transportation Research Record* 2676.8, pp. 363–375. ISSN: 0361-1981. DOI: 10.1177/03611981221083617.

AGATZ, NIELS; ALAN ERERA; MARTIN SAVELSBERGH; XING WANG (Dec. 2012). "Optimization for Dynamic Ride-Sharing: A Review". In: *European Journal of Operational Research* 223.2, pp. 295–303. ISSN: 03772217. DOI: 10.1016/j.ejor.2012.05.028.

ALAM, MD JAHEDUL; MUHAMMAD AHSANUL HABIB (Jan. 2018). "Investigation of the Impacts of Shared Autonomous Vehicle Operation in Halifax, Canada Using a Dynamic Traffic Microsimulation Model". In: *Procedia Computer Science*. The 9th International Conference on Ambient Systems, Networks and Technologies (ANT 2018) / The 8th International Conference on Sustainable Energy Information Technology (SEIT-2018) / Affiliated Workshops 130, pp. 496–503. ISSN: 1877-0509. DOI: 10.1016/j.procs.2018.04.066.

ALAZZAWI, SABINA; MATHIAS HUMMEL; PASCAL KORDT; THORSTEN SICKENBERGER; CHRISTIAN WIESEOTTE; OLIVER WOHAK (June 2018). "Simulating the Impact of Shared, Autonomous Vehicles on Urban Mobility – a Case Study of Milan". In: *EPiC Series in Engineering*. Vol. 2. EasyChair, pp. 94–110. DOI: 10.29007/2n4h.

ALBERT, MARC; CLAUDIO RUCH; EMILIO FRAZZOLI (Aug. 2019). "Imbalance in Mobility-on-Demand Systems: A Stochastic Model and Distributed Control Approach". In: *ACM Transactions on Spatial Algorithms and Systems* 5.2, 13:1–13:22. ISSN: 2374-0353. DOI: 10.1145/3325914.

ALIPOUR, BABAK; LEONARDO TONETTO; ROOZBEH KETABI; AARON YI DING; JÖRG OTT; AHMED HELMY (Nov. 2019). "Where Are You Going Next? A Practical Multi-dimensional Look at Mobility Prediction". In: *Proceedings of the 22nd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*. MSWIM '19. New York, NY, USA: Association for Computing Machinery, pp. 5–12. ISBN: 978-1-4503-6904-6. DOI: 10.1145/3345768.3355923. (Visited on 06/19/2023).

ALONSO-GONZÁLEZ, MARÍA J.; ODED CATS; NIELS VAN OORT; SASCHA HOOGENDOORN-LANSER; SERGE HOOGENDOORN (Aug. 2021). "What Are the Determinants of the Willingness to Share Rides in Pooled On-Demand Services?" In: *Transportation* 48.4, pp. 1733–1765. ISSN: 1572-9435. DOI: 10.1007/s11116-020-10110-2.

ALONSO-MORA, JAVIER; SAMITHA SAMARANAYAKE; ALEX WALLAR; EMILIO FRAZZOLI; DANIELA RUS (Jan. 2017). "On-Demand High-Capacity Ride-Sharing via Dynamic Trip-Vehicle Assignment". In: *Proceedings of the National Academy of Sciences* 114.3, pp. 462–467. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1611675114.

ALONSO-MORA, JAVIER; ALEX WALLAR; DANIELA RUS (Sept. 2017). "Predictive Routing for Autonomous Mobility-on-Demand Systems with Ride-Sharing". In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Vancouver, BC: IEEE, pp. 3583–3590. ISBN: 978-1-5386-2682-5. DOI: 10.1109/IROS.2017.8206203.

ATASOY, BILGE; TAKURO IKEDA; XIANG SONG; MOSHE E. BEN-AKIVA (July 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: 0968-090X. DOI: 10.1016/j.trc.2015.04.009.

AVITAL, MICHEL; MAGNUS ANDERSSON; JEFFREY NICKERSON; ARUN SUNDARARAJAN; MARSHALL VAN ALSTYNE; DEB VERHOEVEN (2014). "The collaborative economy: a disruptive innovation or much ado about nothing?" In: *Proceedings of the 35th International Conference on Information Systems; ICIS 2014*. Association for Information Systems. AIS Electronic Library (AISeL), pp. 1–7.

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 (Jan. 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, 2169-4052. DOI: 10.3141/2564-03.

BASMADJIAN, ROBERT; BENEDIKT KIRPES; JAN MRKOS; MAREK CUCHÝ; SAHAR RASTEGAR (Nov. 2019). "An Interoperable Reservation System for Public Electric Vehicle Charging Stations: A Case Study in Germany". In: *Proceedings of the 1st ACM International Workshop on Technology Enablers and Innovative Applications for Smart Cities and Communities*. New York NY USA: ACM, pp. 22–29. ISBN: 978-1-4503-7015-8. DOI: 10.1145/3364544.3364825.

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 (Dec. 2018). "Automated Mobility-on-Demand vs. Mass Transit: A Multi-Modal Activity-Driven Agent-Based Simulation Approach". In: *Transportation Research Record* 2672.8, pp. 608–618. ISSN: 0361-1981. DOI: 10.1177/0361198118758630.

BAUER, GORDON S.; JEFFERY B. GREENBLATT; BRIAN F. GERKE (Apr. 2018). "Cost, Energy, and Environmental Impact of Automated Electric Taxi Fleets in Manhattan". In: *Environmental Science & Technology* 52.8, pp. 4920–4928. ISSN: 0013-936X. DOI: 10.1021/acs.est.7b04732.

BECKER, FELIX; KAY W. AXHAUSEN (Nov. 2017). "Literature Review on Surveys Investigating the Acceptance of Automated Vehicles". In: *Transportation* 44.6, pp. 1293–1306. ISSN: 1572-9435. DOI: 10.1007/s11116-017-9808-9.

BECKER, HENRIK; FRANCESCO CIARI; KAY W. AXHAUSEN (Mar. 2017). "Comparing Car-Sharing Schemes in Switzerland: User Groups and Usage Patterns". In: *Transportation Research Part A: Policy and Practice* 97, pp. 17–29. ISSN: 09658564. DOI: 10.1016/j.tra.2017.01.004.

BEKTAS, TOLGA; PANAGIOTIS P. REPOUSSIS; CHRISTOS D. TARANTILIS (Nov. 2014). "Chapter 11: Dynamic Vehicle Routing Problems". In: *Vehicle Routing*. Ed. by PAOLO TOTH; DANIELE VIGO. Philadelphia, PA: Society for Industrial and Applied Mathematics, pp. 299–347. ISBN: 978-1-61197-358-7 978-1-61197-359-4. DOI: 10.1137/1.9781611973594.ch11.

BELLAN, REBECCA (Dec. 2022). *Cruise Soft-Launches Robotaxi Rides in Phoenix and Austin*.

BELLMAN, RICHARD (1958). "On a routing problem". In: *Quarterly of applied mathematics* 16.1, pp. 87–90.

BILALI, ALEDIA; FLORIAN DANDL; ULRICH FASTENRATH; KLAUS BOGENBERGER (Oct. 2019). "An Analytical Model for On-Demand Ride Sharing to Evaluate the Impact of Reservation, Detour and Maximum Waiting Time". In: *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 1715–1720. DOI: 10.1109/ITSC.2019.8917280.

BILALI, ALEDIA; ULRICH FASTENRATH; KLAUS BOGENBERGER (Apr. 2022). "Analytical Model to Estimate Ride Pooling Traffic Impacts by Using the Macroscopic Fundamental Diagram". In: *Transportation Research Record* 2676.4, pp. 697–709. ISSN: 0361-1981. DOI: 10.1177/03611981211064892.

BISCHOFF, JOSCHKA; MICHAL MACIEJEWSKI (Jan. 2016). "Autonomous Taxicabs in Berlin – A Spatiotemporal Analysis of Service Performance". In: *Transportation Research Procedia*. Transforming Urban Mobility. Mobil.TUM 2016. International Scientific Conference on Mobility and Transport. Conference Proceedings 19, pp. 176–186. ISSN: 2352-1465. DOI: 10.1016/j.trpro.2016.12.078.

BLUM, PHILIPP; LISA HAMM; ALLISTER LODER; KLAUS BOGENBERGER (2022). "Conceptualizing an Individual Full-Trip Tradable Credit Scheme for Multi-Modal Demand and Supply Management: The MobilityCoin System". In: *Frontiers in Future Transportation* 3. ISSN: 2673-5210.

BOEING, GEOFF (2017). "OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks". In: *Computers, Environment and Urban Systems* 65, pp. 126–139.

BOESCH, PATRICK M.; FRANCESCO CIARI; KAY W. AXHAUSEN (Jan. 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, 2169-4052. DOI: 10.3141/2542-13.

BÖSCH, PATRICK M.; FELIX BECKER; HENRIK BECKER; KAY W. AXHAUSEN (May 2018). "Cost-Based Analysis of Autonomous Mobility Services". In: *Transport Policy* 64, pp. 76–91. ISSN: 0967-070X. DOI: 10.1016/j.tranpol.2017.09.005.

BOTSMAN, RACHEL (2015). "Defining the sharing economy: what is collaborative consumption–and what isn't". In: *Fast Company* 27, p. 2015.

BOTSMAN, RACHEL; ROO ROGERS (2010). "What's mine is yours". In: *The rise of collaborative consumption.*

BURNS, LAWRENCE D; WILLIAM C JORDAN; BONNIE A SCARBOROUGH (2012). "TRANSFORMING PERSONAL MOBILITY". In.

CHAFKIN, MAX; MARK BERGEN (Mar. 2017). "Did Uber Steal the Driverless Future From Google?" In: *Bloomberg.com.*

CHANG, KANG-TSUNG (2006). *Introduction to geographic information systems.* McGraw-Hill Higher Education Boston.

CHANG, KH (2015). "Multiobjective Optimization and Advanced Topics". In: *Design Theory and Methods Using CAD/CAE*, pp. 325–406.

CHEN, T. DONNA; KARA M. KOCKELMAN; JOSIAH P. HANNA (Dec. 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.

CHOUAKI, TAREK; JAKOB PUCHINGER (Jan. 2021). "Agent Based Simulation for the Design of a Mobility Service in the Paris-Saclay Area". In: *Transportation Research Procedia*. 23rd EURO Working Group on Transportation Meeting, EWGT 2020, 16-18 September 2020, Paphos, Cyprus 52, pp. 677–683. ISSN: 2352-1465. DOI: 10.1016/j.trpro.2021.01.081.

CNBC (2018). *Disruptor 50 companies.* https://www.cnbc.com/2018/05/22/meet-the-2018-cnbc-disruptor-50-companies.html. [Online; accessed 3-September-2020].

CNBC (2019). *DiDi Chuxing.* https://www.cnbc.com/2019/05/14/didi-chuxing-2019-disruptor-50.html. [Online; accessed 3-September-2020].

CONFORTI, MICHELE; GÉRARD CORNUÉJOLS; GIACOMO ZAMBELLI, et al. (2014). *Integer programming*. Vol. 271. Springer.

DANDL, FLORIAN (2022). "Operation and Regulation of Autonomous Mobility-on-Demand Systems". PhD thesis. Technische Universität München.

DANDL, FLORIAN; KLAUS BOGENBERGER (Apr. 2018). *Booking Processes in Autonomous Carsharing and Taxi Systems*. Vienna. DOI: 10.5281/zenodo.1451436.

DANDL, FLORIAN; KLAUS BOGENBERGER (June 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: 1558-0016.
DOI: 10.1109/TITS.2018.2857208.

DANDL, FLORIAN; KLAUS BOGENBERGER; HANI S. MAHMASSANI (June 2019). "Autonomous Mobility-on-Demand Real-Time Gaming Framework". In: *2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. Cracow, Poland: IEEE, pp. 1–10. ISBN: 978-1-5386-9484-8. DOI: 10.1109/MTITS.2019.8883286.

DANDL, FLORIAN; BENEDIKT BRACHER; KLAUS BOGENBERGER (June 2017). "Microsimulation of an Autonomous Taxi-System in Munich". In: *2017 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; KLAUS BOGENBERGER (Sept. 2021). "On the Dynamism of User Rejections in Mobility-on-Demand Systems". In: *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 3399–3404.
DOI: 10.1109/ITSC48978.2021.9564918.

DANDL, FLORIAN; FABIAN FEHN; KLAUS BOGENBERGER; FRITZ BUSCH (Nov. 2020). "Pre-Day Scheduling of Charging Processes in Mobility-on-Demand Systems Considering Electricity Price and Vehicle Utilization Forecasts". In: *2020 Forum on Integrated and Sustainable Transportation Systems (FISTS)*, pp. 127–134. DOI: 10.1109/FISTS46898.2020.9264862.

DANDL, FLORIAN; MICHAEL HYLAND; KLAUS BOGENBERGER; HANI S. MAHMASSANI (May 2019). "Evaluating the Impact of Spatio-Temporal Demand Forecast Aggregation on the Operational Performance of Shared Autonomous Mobility Fleets". In: *Transportation*. ISSN: 0049-4488, 1572-9435. 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.

DIA, HUSSEIN; FARID JAVANSHOUR (2017). "Autonomous Shared Mobility-On-Demand: Melbourne Pilot Simulation Study". In: *Transportation Research Procedia* 22, pp. 285–296. ISSN: 23521465. DOI: 10.1016/j.trpro.2017.03.035.

DIJKSTRA, EDSGER WYBE (1959). "A note on two problems in connexion with graphs:(Numerische Mathematik, 1 (1959), p 269-271)". In.

DOCHERTY, IAIN; GREG MARSDEN; JILLIAN ANABLE (2018). "The governance of smart mobility". In: *Transportation Research Part A: Policy and Practice* 115, pp. 114–125.

DUAN, LEYI; YUGUANG WEI; JINCHUAN ZHANG; YANG XIA (Feb. 2020). "Centralized and Decentralized Autonomous Dispatching Strategy for Dynamic Autonomous Taxi Operation in Hybrid Request Mode". In: *Transportation Research Part C: Emerging Technologies* 111, pp. 397–420. ISSN: 0968-090X. DOI: 10.1016/j.trc.2019.12.020.

ENGELHARDT, ROMAN; KLAUS BOGENBERGER (June 2021). "Benefits of Flexible Boarding Locations in On-Demand Ride-Pooling Systems". In: *2021 7th International Conference*

on Models and Technologies for Intelligent Transportation Systems (MT-ITS). Heraklion, Greece: IEEE, pp. 1–6. ISBN: 978-1-72818-995-6. DOI: 10.1109/MT-ITS49943.2021.9529284.

ENGELHARDT, ROMAN; FLORIAN DANDL; ALEDIA BILALI; KLAUS BOGENBERGER (Oct. 2019). "Quantifying the Benefits of Autonomous On-Demand Ride-Pooling: A Simulation Study for Munich, Germany". In: *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. Auckland, New Zealand: IEEE, pp. 2992–2997. ISBN: 978-1-5386-7024-8. DOI: 10.1109/ITSC.2019.8916955. (Visited on 03/13/2020).

ENGELHARDT, ROMAN; FLORIAN DANDL; KLAUS BOGENBERGER (July 2020). *Speed-up Heuristic for an On-Demand Ride-Pooling Algorithm*. DOI: 10.48550/arXiv.2007.14877. arXiv: 2007.14877 [cs, eess, math].

ENGELHARDT, ROMAN; FLORIAN DANDL; KLAUS BOGENBERGER (Oct. 2022). *Simulating Ride-Pooling Services with Pre-Booking and On-Demand Customers*. DOI: 10.48550/arXiv.2210.06972.

ENGELHARDT, ROMAN; FLORIAN DANDL; ARSLAN-ALI SYED; YUNFEI ZHANG; FABIAN FEHN; FYNN WOLF; KLAUS BOGENBERGER (July 2022). *FleetPy: A Modular Open-Source Simulation Tool for Mobility On-Demand Services*. DOI: 10.48550/arXiv.2207.14246. arXiv: 2207.14246 [cs, eess].

ERDMANN, MARVIN; FLORIAN DANDL; KLAUS BOGENBERGER (Apr. 2019). "Dynamic Car-Passenger Matching Based on Tabu Search Using Global Optimization with Time Windows". In: *2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO)*, pp. 1–5. DOI: 10.1109/ICMSAO.2019.8880293.

ERDMANN, MARVIN; FLORIAN DANDL; KLAUS BOGENBERGER (May 2021). "Combining Immediate Customer Responses and Car–Passenger Reassignments in on-Demand Mobility Services". In: *Transportation Research Part C: Emerging Technologies* 126, p. 103104. ISSN: 0968090X. DOI: 10.1016/j.trc.2021.103104.

FAGNANT, DANIEL J.; KARA M. KOCKELMAN (Mar. 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 (Jan. 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, 1572-9435. DOI: 10.1007/s11116-016-9729-z. (Visited on 03/13/2020).

FAGNANT, DANIEL J.; KARA M. KOCKELMAN; PRATEEK BANSAL (Jan. 2016). "Operations of Shared Autonomous Vehicle Fleet for Austin, Texas, Market". In: *Transportation Research Record: Journal of the Transportation Research Board* 2563.1, pp. 98–106. ISSN: 0361-1981, 2169-4052. DOI: 10.3141/2536-12. (Visited on 04/22/2020).

FARHAN, J.; T. DONNA CHEN (Aug. 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; ROMAN ENGELHARDT; KLAUS BOGENBERGER (Sept. 2021). "Ride-Parcel-Pooling - Assessment of the Potential in Combining On-Demand Mobility and City Logistics". In: *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 3366–3372. DOI: 10.1109/ITSC48978.2021.9564630.

FEHN, FABIAN; FLORIAN NOACK; FRITZ BUSCH (June 2019). "Modeling of Mobility On-Demand Fleet Operations Based on Dynamic Electricity Pricing". In: *2019 6th International Confer-*

ence on Models and Technologies for Intelligent Transportation Systems (MT-ITS), pp. 1–6.
DOI: 10.1109/MTITS.2019.8883370.

FIELBAUM, ANDRES; XIAOSHAN BAI; JAVIER ALONSO-MORA (May 2021). "On-Demand Ridesharing with Optimized Pick-up and Drop-off Walking Locations". In: *Transportation Research Part C: Emerging Technologies* 126, p. 103061. ISSN: 0968090X.
DOI: 10.1016/j.trc.2021.103061.

FLOYD, ROBERT W. (June 1962). "Algorithm 97: Shortest Path". In: *Communications of the ACM* 5.6, p. 345. ISSN: 0001-0782. DOI: 10.1145/367766.368168. (Visited on 06/19/2023).

FORD, LR (1956). "Network Flow Theory Paper P-923". In: *Santa Monica, California: RAND Corporation*.

FREI, CHARLOTTE; MICHAEL HYLAND; HANI S. MAHMASSANI (Mar. 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.

FULTON, LEWIS M.; JOAN OGDEN (Feb. 2021). "Sustainable Transportation Energy Pathways". In: *Transportation Research Part D: Transport and Environment* 91, p. 102683.
ISSN: 13619209. DOI: 10.1016/j.trd.2020.102683.

GANSKY, LISA (2010). *The mesh: Why the future of business is sharing*. Penguin.

GENDREAU, M.; G. LAPORTE; F. SEMET (Jan. 2006). "The Maximal Expected Coverage Relocation Problem for Emergency Vehicles". In: *Journal of the Operational Research Society* 57.1, pp. 22–28. ISSN: 1476-9360.
DOI: 10.1057/palgrave.jors.2601991. (Visited on 06/19/2023).

GENDREAU, MICHEL; GILBERT LAPORTE; FRÉDÉRIC SEMET (Nov. 2001). "A Dynamic Model and Parallel Tabu Search Heuristic for Real-Time Ambulance Relocation". In: *Parallel Computing*. Applications of Parallel Computing in Transportation 27.12, pp. 1641–1653. ISSN: 0167-8191. DOI: 10.1016/S0167-8191(01)00103-X. (Visited on 06/19/2023).

GOLPAYEGANI, FATEMEH; MAXIME GUÉRIAU; PIERRE-ANTOINE LAHAROTTE; SAEEDEH GHANADBASHI; JIAYING GUO; JACK GERAGHTY; SHEN WANG (2022). "Intelligent Shared Mobility Systems: A Survey on Whole System Design Requirements, Challenges and Future Direction". In: *IEEE Access* 10, pp. 35302–35320. ISSN: 2169-3536.
DOI: 10.1109/ACCESS.2022.3162848.

GREENBLATT, JEFFERY B.; SAMVEG SAXENA (Sept. 2015). "Autonomous Taxis Could Greatly Reduce Greenhouse-Gas Emissions of US Light-Duty Vehicles". In: *Nature Climate Change* 5.9, pp. 860–863. ISSN: 1758-6798. DOI: 10.1038/nclimate2685.

GREENBLATT, JEFFERY B.; SUSAN SHAHEEN (Sept. 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.

GRODZEVICH, OLEG; OLEKSANDR ROMANKO (Aug. 2006). *Normalization and Other Topics in Multi-Objective Optimization*.

GURUMURTHY, KRISHNA MURTHY; KARA M. KOCKELMAN (Sept. 2018). "Analyzing the Dynamic Ride-Sharing Potential for Shared Autonomous Vehicle Fleets Using Cellphone Data from Orlando, Florida". In: *Computers, Environment and Urban Systems* 71, pp. 177–185.
ISSN: 0198-9715. DOI: 10.1016/j.compenvurbsys.2018.05.008.

GURUMURTHY, KRISHNA MURTHY; KARA M. KOCKELMAN; MICHELE D. SIMONI (June 2019).
"Benefits and Costs of Ride-Sharing in Shared Automated Vehicles across Austin, Texas: Opportunities for Congestion Pricing". In: *Transportation Research Record* 2673.6, pp. 548–556. ISSN: 0361-1981. DOI: 10.1177/0361198119850785.

HAMZEHI, SASCHA; KLAUS BOGENBERGER; PHILIPP FRANECK; BERND KALTENHAUSER (Oct. 2019). "Combinatorial Reinforcement Learning of Linear Assignment Problems". In: *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. Auckland, New Zealand: IEEE, pp. 3314–3321. ISBN: 978-1-5386-7024-8.
DOI: 10.1109/ITSC.2019.8916920. (Visited on 03/14/2020).

HARDT, CORNELIUS; KLAUS BOGENBERGER (Sept. 2020). "From Booking Data to Demand Knowledge Unconstraining Carsharing Demand". In: *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–7.
DOI: 10.1109/ITSC45102.2020.9294413.

HAWKINS, ANDREW J. (Sept. 2021). *Cruise Gets the Green Light to Give Driverless Rides to Passengers in San Francisco.* https://www.theverge.com/2021/9/30/22702962/cruise-waymo-california-dmv-autonomous-vehicle-permit.

HAWKINS, ANDREW J. (Nov. 2022). *Waymo's Driverless Taxis Keep Making Incremental Progress, While Others Flounder.* https://www.theverge.com/2022/11/11/23453262/waymo-av-driverless-taxi-phoenix-california-dmv-progress.

HEILIG, MICHAEL; TIM HILGERT; NICOLAI MALLIG; MARTIN KAGERBAUER; PETER VORTISCH (2017). "Potentials of Autonomous Vehicles in a Changing Private Transportation System – a Case Study in the Stuttgart Region". In: *Transportation Research Procedia* 26, pp. 13–21. ISSN: 23521465. DOI: 10.1016/j.trpro.2017.07.004.

HENAO, ALEJANDRO (2017). *Impacts of Ridesourcing-Lyft and Uber-on Transportation Including VMT, Mode Replacement, Parking, and Travel Behavior.* University of Colorado at Denver.

HENAO, ALEJANDRO; WESLEY E. MARSHALL (Dec. 2019). "The Impact of Ride-Hailing on Vehicle Miles Traveled". In: *Transportation* 46.6, pp. 2173–2194. ISSN: 1572-9435.
DOI: 10.1007/s11116-018-9923-2.

History of Self-Driving Cars Milestones (July 2020). https://www.digitaltrends.com/cars/history-of-self-driving-cars-milestones/.

HO, SIN C.; W.Y. SZETO; YONG-HONG KUO; JANNY M.Y. LEUNG; MATTHEW PETERING; TERENCE W.H. TOU (May 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 (May 2019). "Fleet Operational Policies for Automated Mobility: A Simulation Assessment for Zurich". In: *Transportation Research Part C: Emerging Technologies* 102, pp. 20–31. ISSN: 0968-090X.
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 (Jan. 2018). "A First Look at Bridging Discrete Choice Modeling and Agent-Based Microsimulation in MATSim". In: *Procedia Computer Science.* The 9th International Conference on Ambient Systems, Networks and Technologies (ANT 2018) / The 8th International Conference on Sustainable Energy Information Technology (SEIT-2018) / Affiliated Workshops 130, pp. 900–907. ISSN: 1877-0509.
DOI: 10.1016/j.procs.2018.04.087.

HÖRL, SEBASTIAN; MILOS BALAC; KAY W. AXHAUSEN (June 2019). "Dynamic Demand Estimation for an AMoD System in Paris". In: *2019 IEEE Intelligent Vehicles Symposium (IV)*, pp. 260–266. DOI: 10.1109/IVS.2019.8814051.

HOTTUNG, ANDRé KEVIN TIERNEY (2020). "Neural Large Neighborhood Search for the Capacitated Vehicle Routing Problem". In: *ECAI 2020*. IOS Press, pp. 443–450. DOI: 10.3233/FAIA200124. (Visited on 06/19/2023).

How to Give a Lyft Ride (Feb. 2023). https://www.lyft.com/hub/posts/how-to-give-a-ride.

How Uber Works (Feb. 2023). *How Uber Works for Drivers and Riders — Overview.* https://www.uber.com/us/en/about/how-does-uber-work/.

HYLAND, MICHAEL; HANI S. MAHMASSANI (July 2018). "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 (Apr. 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.

HYLAND, MICHAEL F.; HANI S. MAHMASSANI (Jan. 2017). "Taxonomy of Shared Autonomous Vehicle Fleet Management Problems to Inform Future Transportation Mobility". In: *Transportation Research Record: Journal of the Transportation Research Board* 2653.1, pp. 26–34. ISSN: 0361-1981, 2169-4052. DOI: 10.3141/2653-04.

IACOBUCCI, JOE; KIRK HOVENKOTTER; JACOB ANBINDER (2017). "Transit Systems and the Impacts of Shared Mobility". In: *Disrupting Mobility: Impacts of Sharing Economy and Innovative Transportation on Cities*. Ed. by GEREON MEYER; SUSAN SHAHEEN. Lecture Notes in Mobility. Cham: Springer International Publishing, pp. 65–76. ISBN: 978-3-319-51602-8. DOI: 10.1007/978-3-319-51602-8_4.

IACOBUCCI, RICCARDO; BENJAMIN MCLELLAN; TETSUO TEZUKA (Mar. 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: 0968-090X. DOI: 10.1016/j.trc.2019.01.011.

JÄGER, BENEDIKT; CARSTEN BRICKWEDDE; MARKUS LIENKAMP (Dec. 2018). "Multi-Agent Simulation of a Demand-Responsive Transit System Operated by Autonomous Vehicles". In: *Transportation Research Record* 2672.8, pp. 764–774. ISSN: 0361-1981. DOI: 10.1177/0361198118786644.

JAVANSHOUR, FARID; HUSSEIN DIA; GORDON DUNCAN (Nov. 2019). "Exploring the Performance of Autonomous Mobility On-Demand Systems under Demand Uncertainty". In: *Transportmetrica A: Transport Science* 15.2, pp. 698–721. ISSN: 2324-9935, 2324-9943. DOI: 10.1080/23249935.2018.1528485.

JEANNETTE (June 2020). *The Start of the Autonomous Age at Schiphol Airport, 1997*.

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.

JONES, ERICK C.; BENJAMIN D. LEIBOWICZ (July 2019). "Contributions of Shared Autonomous Vehicles to Climate Change Mitigation". In: *Transportation Research Part D: Transport and Environment* 72, pp. 279–298. ISSN: 1361-9209. DOI: 10.1016/j.trd.2019.05.005.

JUNG, JAEYOUNG; JOSEPH Y. J. CHOW; R. JAYAKRISHNAN; JI YOUNG PARK (Mar. 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: 0968-090X. DOI: 10.1016/j.trc.2014.01.008.

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: 1467-8667. DOI: 10.1111/mice.12157.

Al-Kanj, Lina; Juliana Nascimento; Warren B. Powell (Aug. 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.

Kayikci, Yasanur; Ozgur Kabadurmus (Dec. 2022). "Barriers to the Adoption of the Mobility-as-a-Service Concept: The Case of Istanbul, a Large Emerging Metropolis". In: *Transport Policy* 129, pp. 219–236. issn: 0967-070X. DOI: 10.1016/j.tranpol.2022.10.015.

Kondor, Dániel; Hongmou Zhang; Remi Tachet; Paolo Santi; Carlo Ratti (Aug. 2019). "Estimating Savings in Parking Demand Using Shared Vehicles for Home–Work Commuting". In: *IEEE Transactions on Intelligent Transportation Systems* 20.8, pp. 2903–2912. issn: 1558-0016. DOI: 10.1109/TITS.2018.2869085.

Kong, Hui; Xiaohu Zhang; Jinhua Zhao (June 2020). "How Does Ridesourcing Substitute for Public Transit? A Geospatial Perspective in Chengdu, China". In: *Journal of Transport Geography* 86, p. 102769. issn: 0966-6923. DOI: 10.1016/j.jtrangeo.2020.102769.

Korte, Bernhard H; Jens Vygen (2018). *Combinatorial optimization*. Vol. 1. Springer.

Kucharski, Rafał; Oded Cats (Sept. 2020). "Exact Matching of Attractive Shared Rides (ExMAS) for System-Wide Strategic Evaluations". In: *Transportation Research Part B: Methodological* 139, pp. 285–310. issn: 0191-2615. DOI: 10.1016/j.trb.2020.06.006.

Kuhn, H. W. (1955). "The Hungarian Method for the Assignment Problem". In: *Naval Research Logistics Quarterly* 2.1-2, pp. 83–97. issn: 1931-9193. DOI: 10.1002/nav.3800020109.

Lamotte, Raphaël; André de Palma; Nikolas Geroliminis (May 2017). "On the Use of Reservation-Based Autonomous Vehicles for Demand Management". In: *Transportation Research Part B: Methodological* 99, pp. 205–227. issn: 0191-2615. DOI: 10.1016/j.trb.2017.01.003.

Lazarus, Jessica R.; Juan D. Caicedo; Alexandre M. Bayen; Susan A. Shaheen (June 2021). "To Pool or Not to Pool? Understanding Opportunities, Challenges, and Equity Considerations to Expanding the Market for Pooling". In: *Transportation Research Part A: Policy and Practice* 148, pp. 199–222. issn: 09658564. DOI: 10.1016/j.tra.2020.10.007.

Levin, Michael W. (Sept. 2017). "Congestion-Aware System Optimal Route Choice for Shared Autonomous Vehicles". In: *Transportation Research Part C: Emerging Technologies* 82, pp. 229–247. issn: 0968-090X. DOI: 10.1016/j.trc.2017.06.020.

Lewis, Elyse O'C.; Don MacKenzie (Jan. 2017). "UberHOP in Seattle: Who, Why, and How?" In: *Transportation Research Record* 2650.1, pp. 101–111. issn: 0361-1981. DOI: 10.3141/2650-12. (Visited on 06/19/2023).

Li, Dun; Youlin Huang; Lixian Qian (Sept. 2022). "Potential Adoption of Robotaxi Service: The Roles of Perceived Benefits to Multiple Stakeholders and Environmental Awareness". In: *Transport Policy* 126, pp. 120–135. issn: 0967-070X. DOI: 10.1016/j.tranpol.2022.07.004.

Li, Li; DianChao Lin; Theodoros Pantelidis; Joseph Chow; Saif Eddin Jabari (Oct. 2019). "An Agent-based Simulation for Shared Automated Electric Vehicles with Vehicle Relocation". In: *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 3308–3313. DOI: 10.1109/ITSC.2019.8917253.

Li, Li; Theodoros Pantelidis; Joseph Y. J. Chow; Saif Eddin Jabari (Aug. 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: 1366-5545. DOI: 10.1016/j.tre.2021.102392.

LI, ZIRU; YILI HONG; ZHONGJU ZHANG (Sept. 2016). *An Empirical Analysis of On-Demand Ride Sharing and Traffic Congestion*. SSRN Scholarly Paper. Rochester, NY. (Visited on 06/19/2023).

LIIMATAINEN, HEIKKI; MILOŠ N. MLADENOVIĆ (June 2021). "Developing Mobility as a Service – User, Operator and Governance Perspectives". In: *European Transport Research Review* 13.1, p. 37. ISSN: 1866-8887. DOI: 10.1186/s12544-021-00496-0.

LITMAN, TODD (Jan. 2023). "Autonomous Vehicle Implementation Predictions: Implications for Transport Planning". In.

LIU, JUN; STEVEN JONES; EMMANUEL KOFI ADANU (Jan. 2020). "Challenging Human Driver Taxis with Shared Autonomous Vehicles: A Case Study of Chicago". In: *Transportation Letters* 12.10, pp. 701–705. ISSN: 1942-7867. DOI: 10.1080/19427867.2019.1694202.

LIU, ZHIGUANG; TOMIO MIWA; WEILIANG ZENG; TAKAYUKI MORIKAWA (2018). *An Agent-Based Simulation Model for Shared Autonomous Taxi System*. DOI: 10.11175/eastsats.5.1.

LOEB, BENJAMIN; KARA M. KOCKELMAN; JUN LIU (Apr. 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.

LOKHANDWALA, MUSTAFA; HUA CAI (Dec. 2018). "Dynamic Ride Sharing Using Traditional Taxis and Shared Autonomous Taxis: A Case Study of NYC". In: *Transportation Research Part C: Emerging Technologies* 97, pp. 45–60. ISSN: 0968090X. DOI: 10.1016/j.trc.2018.10.007.

LYU, TAO; PEIRONG (SLADE) WANG; YANAN GAO; YUANQING WANG (Feb. 2021). "Research on the Big Data of Traditional Taxi and Online Car-Hailing: A Systematic Review". In: *Journal of Traffic and Transportation Engineering (English Edition)* 8.1, pp. 1–34. ISSN: 2095-7564. DOI: 10.1016/j.jtte.2021.01.001.

MA, JIAQI; XIAOPENG LI; FANG ZHOU; WEI HAO (Nov. 2017). "Designing Optimal Autonomous Vehicle Sharing and Reservation Systems: A Linear Programming Approach". In: *Transportation Research Part C: Emerging Technologies* 84, pp. 124–141. ISSN: 0968090X. DOI: 10.1016/j.trc.2017.08.022.

MACIEJEWSKI, MICHAL; JOSCHKA BISCHOFF (Dec. 2018). "CONGESTION EFFECTS OF AUTONOMOUS TAXI FLEETS". In: *Transport* 33.4, pp. 971–980. ISSN: 1648-4142, 1648-3480. DOI: 10.3846/16484142.2017.1347827.

MARASCO, MELISSA (Feb. 2023). *The Future on the Coastside? Zoox the Purpose-Built Robotaxi Is First in the World to Operate on Public Roads*. https://www.coastsidebuzz.com/the-future-on-the-coastside-zoox-the-purpose-built-robotaxi-is-first-in-the-world-to-operate-on-public-roads/.

MARTINEZ, LUIS M.; JOSÉ MANUEL VIEGAS (June 2017). "Assessing the Impacts of Deploying a Shared Self-Driving Urban Mobility System: An Agent-Based Model Applied to the City of Lisbon, Portugal". In: *International Journal of Transportation Science and Technology* 6.1, pp. 13–27. ISSN: 20460430. DOI: 10.1016/j.ijtst.2017.05.005.

MATOUSEK, JIRI; BERND GÄRTNER (2007). *Understanding and using linear programming*. Springer Science & Business Media.

MOAVENZADEH, JOHN; NIKOLAUS S. LANG (2018). *Reshaping Urban Mobility with Autonomous Vehicles: Lessons from the City of Boston - Urbanism Next*. Tech. rep. World Economic Forum.

MOORTHY, ADITI; ROBERT DE KLEINE; GREGORY KEOLEIAN; JEREMY GOOD; GEOFF LEWIS (Mar. 2017). "Shared Autonomous Vehicles as a Sustainable Solution to the Last Mile Problem: A Case Study of Ann Arbor-Detroit Area". In: *SAE International Journal of Passenger Cars - Electronic and Electrical Systems* 10.2, pp. 328–336. ISSN: 1946-4622. DOI: 10.4271/2017-01-1276.

MORENO, JOHAN (Jan. 2021). *Waymo CEO Says Tesla Is Not A Competitor, Gives Estimated Cost Of Autonomous Vehicles*. https://www.forbes.com/sites/johanmoreno/2021/01/22/waymo-ceo-says-tesla-is-not-a-competitor-gives-estimated-cost-of-autonomous-vehicles/.

MURRAY, ALAN T. (Oct. 2003). "A Coverage Model for Improving Public Transit System Accessibility and Expanding Access". In: *Annals of Operations Research* 123.1, pp. 143–156. ISSN: 1572-9338. DOI: 10.1023/A:1026123329433.

NARAYANAN, SANTHANAKRISHNAN; CONSTANTINOS ANTONIOU (Feb. 2023). "Shared Mobility Services towards Mobility as a Service (MaaS): What, Who and When?" In: *Transportation Research Part A: Policy and Practice* 168, p. 103581. ISSN: 0965-8564. DOI: 10.1016/j.tra.2023.103581.

NARAYANAN, SANTHANAKRISHNAN; EMMANOUIL CHANIOTAKIS; CONSTANTINOS ANTONIOU (Feb. 2020). "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.

NAZARI, MOHAMMADREZA; AFSHIN OROOJLOOY; MARTIN TAKÁČ; LAWRENCE V. SNYDER (Dec. 2018). "Reinforcement Learning for Solving the Vehicle Routing Problem". In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. NIPS'18. Red Hook, NY, USA: Curran Associates Inc., pp. 9861–9871. DOI: 10.5555/3327546.3327651. (Visited on 06/19/2023).

NGUYEN-PHUOC, DUY Q.; MENG ZHOU; MING HONG CHUA; ANDRÉ ROMANO ALHO; SIMON OH; RAVI SESHADRI; DIEM-TRINH LE (Mar. 2023). "Examining the Effects of Automated Mobility-on-Demand Services on Public Transport Systems Using an Agent-Based Simulation Approach". In: *Transportation Research Part A: Policy and Practice* 169, p. 103583. ISSN: 0965-8564. DOI: 10.1016/j.tra.2023.103583.

NOCEDAL, JORGE; STEPHEN WRIGHT (2006). *Numerical optimization*. Springer Science & Business Media.

NORDHOFF, SINA; JOOST DE WINTER; MILTOS KYRIAKIDIS; BART VAN AREM; RIENDER HAPPEE (Apr. 2018). "Acceptance of Driverless Vehicles: Results from a Large Cross-National Questionnaire Study". In: *Journal of Advanced Transportation* 2018, e5382192. ISSN: 0197-6729. DOI: 10.1155/2018/5382192.

OPENSTREETMAP (2017). *Planet dump retrieved from https://planet.osm.org*. https://www.openstreetmap.org.

PADBERG, MANFRED; GIOVANNI RINALDI (1987). "Optimization of a 532-city symmetric traveling salesman problem by branch and cut". In: *Operations research letters* 6.1, pp. 1–7.

PANTELIDIS, THEODOROS P.; LI LI; TAI-YU MA; JOSEPH Y. J. CHOW; SAIF EDDIN G. JABARI (May 2022). "A Node-Charge Graph-Based Online Carshare Rebalancing Policy with Capacitated Electric Charging". In: *Transportation Science* 56.3, pp. 654–676. ISSN: 0041-1655. DOI: 10.1287/trsc.2021.1058.

PARENT, M.; P. DAVIET (July 1993). "Automatic Driving For Small Public Urban Vehicles". In: *Proceedings of the Intelligent Vehicles '93 Symposium*, pp. 402–407. DOI: 10.1109/IVS.1993.697360.

PARENT, MICHEL; ARNAUD DE LA FORTELLE (Nov. 2005). "Cybercars : Past, Present and Future of the Technology". In.

PARZEN, EMANUEL (1962). "On Estimation of a Probability Density Function and Mode". In: *The Annals of Mathematical Statistics* 33.3, pp. 1065–1076. ISSN: 0003-4851. JSTOR: 2237880. (Visited on 06/20/2023).

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. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 399–416. ISBN: 978-3-662-45853-2 978-3-662-45854-9. DOI: 10.1007/978-3-662-45854-9_19.

PAVONE, MARCO; STEPHEN L SMITH; EMILIO FRAZZOLI; DANIELA RUS (June 2012). "Robotic Load Balancing for Mobility-on-Demand Systems". In: *The International Journal of Robotics Research* 31.7, pp. 839–854. ISSN: 0278-3649, 1741-3176. DOI: 10.1177/0278364912444766.

PINTO, HELEN K. R. F.; MICHAEL F. HYLAND; HANI S. MAHMASSANI; I. ÖMER VERBAS (Apr. 2020). "Joint Design of Multimodal Transit Networks and Shared Autonomous Mobility Fleets". In: *Transportation Research Part C: Emerging Technologies*. ISTTT 23 TR_C-23rd International Symposium on Transportation and Traffic Theory (ISTTT 23) 113, pp. 2–20. ISSN: 0968-090X. DOI: 10.1016/j.trc.2019.06.010.

POWELL, WARREN B.; YOSEF SHEFFI; KENNETH S. NICKERSON; KEVIN BUTTERBAUGH; SUSAN ATHERTON (Feb. 1988). "Maximizing Profits for North American Van Lines' Truckload Division: A New Framework for Pricing and Operations". In: *Interfaces* 18.1, pp. 21–41. ISSN: 0092-2102, 1526-551X. DOI: 10.1287/inte.18.1.21. (Visited on 09/28/2020).

QIAO, SI; ANTHONY GAR-ON YEH (Jan. 2023). "Is Ride-Hailing Competing or Complementing Public Transport? A Perspective from Affordability". In: *Transportation Research Part D: Transport and Environment* 114, p. 103533. ISSN: 1361-9209. DOI: 10.1016/j.trd.2022.103533.

RAYLE, LISA; DANIELLE DAI; NELSON CHAN; ROBERT CERVERO; SUSAN SHAHEEN (Jan. 2016). "Just a Better Taxi? A Survey-Based Comparison of Taxis, Transit, and Ridesourcing Services in San Francisco". In: *Transport Policy* 45, pp. 168–178. ISSN: 0967-070X. DOI: 10.1016/j.tranpol.2015.10.004.

RODIER, CAROLINE; JULIA MICHAELS (Feb. 2019). "The Effects of Ride-Hailing Services on Greenhouse Gas Emissions". In: DOI: 10.7922/G23T9FDP.

ROTH, EMMA (Nov. 2022). *Waymo Can Now Provide Fully Driverless Rides in San Francisco*. https://www.theverge.com/2022/11/19/23467784/waymo-provide-fully-driverless-rides-san-francisco-california.

RUCH, CLAUDIO; SEBASTIAN HÖRL; EMILIO FRAZZOLI (Nov. 2018). "AMoDeus, a Simulation-Based Testbed for Autonomous Mobility-on-Demand Systems". In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 3639–3644. DOI: 10.1109/ITSC.2018.8569961.

SALAZAR, MAURO; FEDERICO ROSSI; MAXIMILIAN SCHIFFER; CHRISTOPHER H. ONDER; MARCO PAVONE (Nov. 2018). "On the Interaction between Autonomous Mobility-on-Demand and Public Transportation Systems". In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2262–2269. DOI: 10.1109/ITSC.2018.8569381.

SANTI, PAOLO; GIOVANNI RESTA; MICHAEL SZELL; STANISLAV SOBOLEVSKY; STEVEN H. STROGATZ; CARLO RATTI (Sept. 2014). "Quantifying the Benefits of Vehicle Pooling with

Shareability Networks". In: *Proceedings of the National Academy of Sciences* 111.37, pp. 13290–13294. DOI: 10.1073/pnas.1403657111.

Santos, Douglas O.; Eduardo C. Xavier (Nov. 2015). "Taxi and Ride Sharing: A Dynamic Dial-a-Ride Problem with Money as an Incentive". In: *Expert Systems with Applications* 42.19, pp. 6728–6737. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2015.04.060.

Schaller, Bruce (2018a). "The new automobility: Lyft, Uber and the future of American cities". In.

Schaller, Bruce (2018b). "Unsustainable? The Growth of App-Based Ride Services and Traffic, Travel and the Future of New York City". In.

SF Officials Describe Street Chaos From Cruise, Waymo Robotaxis (Jan. 2023). https://sfstandard.com/transportation/sf-officials-describe-chaos-from-cruise-waymo-cars-as-they-try-to-slow-their-rollout/.

Shaheen, Susan; Nelson Chan; Apaar Bansal; Adam Cohen (Nov. 2015). "Shared Mobility: A Sustainability & Technologies Workshop: Definitions, Industry Developments, and Early Understanding". In.

Shaheen, Susan; Adam Cohen (July 2019). "Shared Ride Services in North America: Definitions, Impacts, and the Future of Pooling". In: *Transport Reviews* 39.4, pp. 427–442. ISSN: 0144-1647. DOI: 10.1080/01441647.2018.1497728.

Shaheen, Susan; Adam Cohen; Mark Jaffee (Apr. 2018). "Innovative Mobility: Carsharing Outlook". In: DOI: 10.7922/G2CC0XVW.

Shaheen, Susan A.; Adam P. Cohen; Jacquelyn Broader; Richard Davis; Les Brown; Radha Neelakantan; Deepak Gopalakrishna; ICF; Berkeley. Transportation Sustainability Research Center University of California (Mar. 2020). *Mobility on Demand Planning and Implementation: Current Practices, Innovations, and Emerging Mobility Futures.* Tech. rep. FHWA-JPO-20-792.

Shen, Yu; Hongmou Zhang; Jinhua Zhao (July 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.

Silverman, Bernhard W (1982). "Algorithm AS 176: Kernel density estimation using the fast Fourier transform". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 31.1, pp. 93–99.

Snyman, Jan A; Daniel N Wilke (2018). *Practical Mathematical Optimization: Basic Optimization Theory and Gradient-Based Algorithms.* Vol. 133. Springer.

Song, Jialin; Ravi Lanka; Yisong Yue; Bistra Dilkina (2020). "A General Large Neighborhood Search Framework for Solving Integer Programs". In: *arXiv preprint arXiv:2004.00422.*

Spieser, Kevin; Samitha Samaranayake; Wolfgang Gruel; Emilio Frazzoli (2016). "Shared-Vehicle Mobility-on-Demand Systems: A Fleet Operator's Guide to Rebalancing Empty Vehicles". In: *Transportation Research Board 95th Annual MeetingTransportation Research Board.* 16-5987.

Stocker, Adam; Susan Shaheen (2017). *Shared Automated Vehicles: Review of Business Models.* Working Paper 2017-09. International Transport Forum Discussion Paper.

Stocker, Adam; Susan Shaheen (2019). "Shared Automated Vehicle (SAV) Pilots and Automated Vehicle Policy in the U.S.: Current and Future Developments". In: *Road Vehicle Automation 5.* Ed. by Gereon Meyer; Sven Beiker. Lecture Notes in Mobility. Cham: Springer International Publishing, pp. 131–147. ISBN: 978-3-319-94896-6. DOI: 10.1007/978-3-319-94896-6_12.

SUN, YANSHUO; LEI ZHANG (2018). "Potential of Taxi-Pooling to Reduce Vehicle Miles Traveled in Washington, DC". In: *Transportation Research Record* 2672.8, pp. 775–784.

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. Cham: Springer International Publishing, pp. 584–595. ISBN: 978-3-030-30240-5 978-3-030-30241-2. DOI: 10.1007/978-3-030-30241-2_49.

SYED, ARSLAN ALI; FLORIAN DANDL; KLAUS BOGENBERGER (Sept. 2021). "User-Assignment Strategy Considering Future Imbalance Impacts for Ride Hailing". In: *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. Indianapolis, IN, USA: IEEE, pp. 2441–2446. ISBN: 978-1-72819-142-3. DOI: 10.1109/ITSC48978.2021.9564559.

SYED, ARSLAN ALI; FLORIAN DANDL; BERND KALTENHÄUSER; KLAUS BOGENBERGER (June 2021). "Density Based Distribution Model for Repositioning Strategies of Ride Hailing Services". In: *Frontiers in Future Transportation* 2, p. 681451. ISSN: 2673-5210. DOI: 10.3389/ffutr.2021.681451.

SYED, ARSLAN ALI; MARKUS FISCHER; CORNELIUS HARDT; KLAUS BOGENBERGER (Oct. 2022). "Charge Point Search Policies for EVs - Minimizing The Cruise for Juice". In: *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2487–2493. DOI: 10.1109/ITSC55140.2022.9921802.

SYED, ARSLAN ALI; IRINA GAPONOVA; KLAUS BOGENBERGER (May 2019). "Neural Network-Based Metaheuristic Parameterization with Application to the Vehicle Matching Problem in Ride-Hailing Services". In: *Transportation Research Record: Journal of the Transportation Research Board*. ISSN: 0361-1981, 2169-4052. DOI: 10.1177/0361198119846099.

SYED, ARSLAN ALI; BERND KALTENHAEUSER; IRINA GAPONOVA; KLAUS BOGENBERGER (Oct. 2019). "Asynchronous Adaptive Large Neighborhood Search Algorithm for Dynamic Matching Problem in Ride Hailing Services". In: *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. Auckland, New Zealand: IEEE, pp. 3006–3012. ISBN: 978-1-5386-7024-8. DOI: 10.1109/ITSC.2019.8916943.

SYED, ARSLAN ALI; YUNFEI ZHANG; KLAUS BOGENBERGER (Sept. 2023). "Data-Driven Spatio-Temporal Scaling of Travel Times for AMoD Simulations". In: *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. Bilbao, Spain: IEEE, pp. 3583–3588. ISBN: 9798350399462. DOI: 10.1109/ITSC57777.2023.10422313.

Automation Levels-SAE International (Apr. 2021). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles - SAE International.* https://www.sae.org/standards/content/j3016_202104/.

The DARPA Grand Challenge: Ten Years Later (2014). https://www.darpa.mil/news-events/2014-03-13.

TIRACHINI, ALEJANDRO (2019). "Ride-hailing, travel behaviour and sustainable mobility: an international review". In: *Transportation*, pp. 1–37.

TLC Trip Record Data - TLC (2023). https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page.

TONETTO, LEONARDO; MORITZ UNTERSPERGER; JÖRG OTT (2019). "Towards Exploiting Wi-Fi Signals From Low Density Infrastructure for Crowd Estimation". In: *Proceedings of the 14th Workshop on Challenged Networks*, pp. 27–32.

TOTH, PAOLO; DANIELE VIGO (2014). *Vehicle routing: problems, methods, and applications.* SIAM.

TRANSIT, MASS (2019). https://www.masstransitmag.com/alt-mobility/article/21095130/best-practices-tnc-integration. [Online; accessed 3-September-2020].

TURAN, BERKAY; RAMTIN PEDARSANI; MAHNOOSH ALIZADEH (Dec. 2020). "Dynamic Pricing and Fleet Management for Electric Autonomous Mobility on Demand Systems". In: *Transportation Research Part C: Emerging Technologies* 121, p. 102829. ISSN: 0968-090X. DOI: 10.1016/j.trc.2020.102829.

UBER (2020). *Uber cities.* https://www.uber.com/global/en/cities. [Online; accessed 3-September-2020].

UBER (2022). *Uber Matching.* https://www.uber.com/us/en/marketplace/matching/. [Online; accessed 9-September-2022].

Urban Mobility System Upgrade (Mar. 2015). *Urban Mobility System Upgrade: How Shared Self-Driving Cars Could Change City Traffic.* International Transport Forum Policy Papers 6. DOI: 10.1787/5jlwvzdk29g5-en.

VAN HEMERT, JANO I; JOHANNES A LA POUTRÉ (2004). "Dynamic routing problems with fruitful regions: Models and evolutionary computation". In: *International Conference on Parallel Problem Solving from Nature.* Springer, pp. 692–701.

VOSOOGHI, REZA; JAKOB PUCHINGER; JOSCHKA BISCHOFF; MARIJA JANKOVIC; ANTHONY VOUILLON (Apr. 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: 1361-9209. DOI: 10.1016/j.trd.2020.102283.

WALLAR, ALEX; MENNO VAN DER ZEE; JAVIER ALONSO-MORA; DANIELA RUS (2018). "Vehicle rebalancing for mobility-on-demand systems with ride-sharing". In: *2018 IEEE/RSJ international conference on intelligent robots and systems (IROS).* IEEE, pp. 4539–4546.

WANG, DI; TOMIO MIWA; TAKAYUKI MORIKAWA (Oct. 2021). "Comparative Analysis of Spatial–Temporal Distribution between Traditional Taxi Service and Emerging Ride-Hailing". In: *ISPRS International Journal of Geo-Information* 10.10, p. 690. ISSN: 2220-9964. DOI: 10.3390/ijgi10100690.

WANG, HAO; RUEY CHEU; DER-HORNG LEE (Sept. 2014). "Intelligent Taxi Dispatch System for Advance Reservations". In: *Journal of Public Transportation* 17.3. ISSN: 1077-291X. DOI: 10.5038/2375-0901.17.3.8.

WANG, SENLEI; GONÇALO HOMEM DE ALMEIDA CORREIA; HAI XIANG LIN (Nov. 2022). "Modeling the Competition between Multiple Automated Mobility On-Demand Operators: An Agent-Based Approach". In: *Physica A: Statistical Mechanics and its Applications* 605, p. 128033. ISSN: 0378-4371. DOI: 10.1016/j.physa.2022.128033.

WANG, XIANFU (2005). "Volumes of generalized unit balls". In: *Mathematics Magazine* 78.5, pp. 390–395.

WANG, YANSHENG; YONGXIN TONG; CHENG LONG; PAN XU; KE XU; WEIFENG LV (2019). "Adaptive dynamic bipartite graph matching: A reinforcement learning approach". In: *2019 IEEE 35th International Conference on Data Engineering (ICDE).* IEEE, pp. 1478–1489.

WARSHALL, STEPHEN (1962). "A theorem on boolean matrices". In: *Journal of the ACM (JACM)* 9.1, pp. 11–12.

WEDIA (Nov. 2021). *The Future Is Now: Robot Taxis Are Coming to Germany.* https://www.iamexpat.de/expat-info/german-expat-news/future-now-robot-taxis-are-coming-germany.

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 (Aug. 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 (Dec. 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.

WILKES, GABRIEL; ROMAN ENGELHARDT; LARS BRIEM; FLORIAN DANDL; PETER VORTISCH; KLAUS BOGENBERGER; MARTIN KAGERBAUER (Aug. 2021). "Self-Regulating Demand and Supply Equilibrium in Joint Simulation of Travel Demand and a Ride-Pooling Service". In: *Transportation Research Record* 2675.8, pp. 226–239. ISSN: 0361-1981. DOI: 10.1177/0361198121997140.

WINTER, KONSTANZE; ODED CATS; KAREL MARTENS; BART VAN AREM (Jan. 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: 0965-8564. DOI: 10.1016/j.tra.2020.11.008.

WOCKATZ, P; P SCHARTAU (2015). "IM traveller needs and UK capability study: Supporting the realisation of intelligent mobility in the UK". In: *Transport Systems Catapult, Milton Keynes*.

WOLBERTUS, RICK; ROBERT VAN DEN HOED; MAARTEN KROESEN; CASPAR CHORUS (June 2021). "Charging Infrastructure Roll-out Strategies for Large Scale Introduction of Electric Vehicles in Urban Areas: An Agent-Based Simulation Study". In: *Transportation Research Part A: Policy and Practice* 148, pp. 262–285. ISSN: 09658564. DOI: 10.1016/j.tra.2021.04.010.

WOOLDRIDGE, M. (Jan. 1997). "Agent-Based Software Engineering". In: *IEE Proceedings - Software* 144.1, pp. 26–37. ISSN: 1463-9831. DOI: 10.1049/ip-sen:19971026.

ZGRAGGEN, JANNIK; MATTHEW TSAO; MAURO SALAZAR; MAXIMILIAN SCHIFFER; MARCO PAVONE (Oct. 2019). "A Model Predictive Control Scheme for Intermodal Autonomous Mobility-on-Demand". In: *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 1953–1960. DOI: 10.1109/ITSC.2019.8917521.

ZHANG, RICK; MARCO PAVONE (Jan. 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, 1741-3176. DOI: 10.1177/0278364915581863.

ZHANG, RICK; FEDERICO ROSSI; MARCO PAVONE (May 2016). "Model Predictive Control of Autonomous Mobility-on-Demand Systems". In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1382–1389. DOI: 10.1109/ICRA.2016.7487272. arXiv: 1509.03985 [cs].

ZHANG, WENWEN; SUBHRAJIT GUHATHAKURTA (Jan. 2017). "Parking Spaces in the Age of Shared Autonomous Vehicles: How Much Parking Will We Need and Where?" In: *Transportation Research Record* 2651.1, pp. 80–91. ISSN: 0361-1981. DOI: 10.3141/2651-09.

ZHANG, WENWEN; SUBHRAJIT GUHATHAKURTA; JINQI FANG; GE ZHANG (Dec. 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.

ZHANG, YUNFEI; ROMAN ENGELHARDT; ARSLAN-ALI SYED; FLORIAN DANDL; CORNELIUS HARDT; KLAUS BOGENBERGER (Oct. 2022). "Simulating Charging Processes of Mobility-

On-Demand Services at Public Infrastructure: Can Operators Complement Each Other?" In: *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2200–2205. DOI: 10.1109/ITSC55140.2022.9922449.

Zwick, Felix; Kay W. Axhausen (Sept. 2020). "Impact of Service Design on Urban Ridepooling Systems". In: *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. Rhodes, Greece: IEEE, pp. 1–6. ISBN: 978-1-72814-149-7. DOI: 10.1109/ITSC45102.2020.9294289.

Appendix

A. Overview Table for the Literature Reviewed

Section 2.5 discussed the AMoD literature using three main categories: service type (section 2.5.2), overall System modeling (section 2.6), and fleet management (section 2.5.4). This section overviews the AMoD literature included in the dissertation using these categories. For a detailed description of each of the categories, the reader is referred to the corresponding sections mentioned above. For conciseness, the following abbreviations are adopted for each of the categories. If it is clear that the reviewed work must have an entry for a particular category but the description in the paper is unclear, then the dissertation uses the symbol "X" (unknown) for that category. Additionally, to reduce the number of symbols, instead of a separate category "mix" for a combination of possible options for a characteristic, the table directly lists the applicable options.

Service Type

- **Sharing System**: Ride-Hailing (**RH**), Ride-Sharing (**RS**)

- **Booking Type**:
  - Announcement: On-demand(**O**), Reservation based (**RB**)
  - Serving Flexibility: All binding (**AB**), Immediate Rejection (**IR**), Rejection after a period (**RT**)

- **Electrification**: Charging Infrastructure (CI):
  - Allowed to charge at public CI (**PCI**)
  - Specific CI for the AMoD service (**SCI**)

- **Fleet Composition**: Homogeneous (**H**), Non-homogeneous(**NH**)

- **Integration** with other transport modes:
  - Integrated: First and last mile (**FL**), Full network integration (**FN**)
  - Independent (**IND**)

- **Pricing** (includes the objective function used for vehicle assignments)
  - Price structure: base-fare (**BF**), distance-based (**DB**), time-based (**TB**), price split for ride sharing (**SRS**)
  - Dynamism: Surge Pricing (**SP**), Congestion Pricing (**CP**), Detour-based (**DeB**)

AMoD System Model

- **Agent Models**:

- – Customer (**C**): Simple with only request time (**SR**), Leave system based on waiting time (**W**), Accept/Reject (**AR**)
- – Vehicle (**V**): Independent points (**IP**), Interactive with other vehicles (**ItV**)

- **Demand Modeling**: Taxi or MoD data (**Data**), Demand Models (**Model**)

- **Transport Network**:
  - – 2D continuous plane (**CP**)
  - – Grid Network (**GN**)
  - – Node-link network (**NL**):
    - ∗ Coarse (**C**)
    - ∗ Road-level (**R**)

- **Traffic Model**: Manhattan Metric (**MM**), Euclidean Metric (**EM**), Scaled free-flow (**SFF**), Traffic Simulation (**TS**).

- **Integration model** with other transport modes:
  - – Line-based (**LB**)
  - – Traffic Simulation (**TS**)

Fleet Management

- **Vehicle Assignment**:
  - – Dynamism: Immediate Response (**IR**), Batching (**B**)
  - – Solution Method: Heuristic(**H**), Optimization (**O**)
  - – Vehicles considered in the assignment method: Only idle (**I**), Idle and enroute (**E**)
  - – Reassignment (**RA**)
  - – Re-optimization (**RE**)

- **Repositioning**:
  - – Short-term (**ST**):
    - ∗ Explicit repositioning (**E**)
    - ∗ Predictive Routing (**P**)
  - – Mid-term (**MT**)

- **Charging method**: Heuristic (**H**), Dynamic programming (**DP**), Model predictive control (**MPC**), Rest of the optimization-based formulation (**O**),

		Service Type						System Model						Fleet Management					
Citation	City	Sharing System	Booking Type	Electrification CI	Fleet Composition	Integration Type	Pricing	Agent Models	Demand Modeling	Transport Network	Traffic Model	TS Software	Integration Model	Dynamism	Solution Method	Vehicles Considered	Reassign, Re-optimize	Repositioning	Charging Method
Alam et al. [2018]	Halifax	RH	O, RT		H		TB	C (W), V (ItV)	Model	NL (R)	TS	VISSIM		IR	H	I			
Alazzawi et al. [2018]	Milan	RH	O, AB		H		TB	C (SR), V(ItV)	Model	NL (R)	TS	SUMO		IR	H	I			
Al-Kanj et al. [2020]	Virtual	RH	O, IR	SCI	H		DB, SP	C(W,AR), V(IP)	Model	GN				B	O	I, E			DP
Alonso-Mora et al. [2017]	Manhattan	RS	O, IR		H		TB, DeB	C(W), V(IP)	Data	NL (R)	SFF			B	O	I, E	RA, RE	ST (E)	
Alonso-Mora et al. [2017]	Manhattan	RS	O, IR		H		TB, DeB	C(W), V(IP)	Data	NL (R)	SFF			B	O	I, E	RA, RE	ST (P)	
Atasoy et al. [2015]	Hino	RS	O, RB, IR		NH		BF,TB, DB	C(W,AR), V(IP)	Model	NL (R)	SFF			B	O	I, E			
Azevedo et al. [2016]	Singapore	RH	O, AB		H	IND	TB,DB	C(SR), V(ItV)	Model	NL (R)	TS	SimMobility	TS	B	H, O	I, E		MT	
Basu et al. [2018]	Virtual	RS	O, AB		H	FL	TB,DB	C(SR), V(ItV)	Model	NL (R)	TS	SimMobility	TS	B	O	I, E		MT	
Bauer et al. [2018]	Manhattan	RH	O, AB	SCI	H		DB	C(SR), V(IP)	Data	NL (R)	SFF			B	H	I		MT	H
Bilali et al. [2022]	Munich	RS	O, AB		H		DB	C(SR), V(ItV)	Data	NL (R)	TS	Aimsun		B	O	I, E			
Bischoff et al. [2016]	Berlin	RH	O, AB		H		DB	C(SR), V(ItV)	Model	NL (R)	TS	MATSIM		B	H	I			
Boesch et al. [2016]	Zurich	RH	O, RT		H		DB	C(W), V(ItV)	Model	NL (R)	TS	MATSIM		IR	H	I			
Chen et al. [2016]	Austin	RH	O, AB	SCI	H		DB	C(SR), V(IP)	Model	GN	MM			IR	H	I		ST (E)	H
Chouaki et al. [2021]	Paris	RS	O, AB	SCI	H	IND	TB, DB	C(SR), V(ItV)	Model	NL (R)	TS	SUMO	TS	IR	H	I, E			H
Dandl et al. [2017]	Munich	RH	O, IR		H		TB	C(W,AR), V(ItV)	Model	NL (R)	TS	Aimsun		IR	H	I, E			
Dandl et al. [2018]	Munich	RH	O, IR		H		TB	C(W), V(IP)	Model	NL (R)	SFF			IR, B	O	I, E	RA, RE		
Dandl et al. [2019]	Munich	RH	O, IR	SCI	H		TB, DB	C(W), V(ItV)	Data	NL (R)	SFF			B	O	I, E		MT	H
Dandl et al. [2019]	Manhattan	RH	O,IR, RT		H		TB, DB	C(SR), V(IP)	Data	NL (R)	SFF			IR, B	H, O	I, E	RA, RE		
Dandl et al. [2019]	Manhattan	RH	O, AB		H		TB, DB, BF	C(SR), V(IP)	Data	NL (R)	SFF			B	O	I, E		ST (E)	
Dandl et al. [2020]	Manhattan, Chicago	RH	O, IR		H		BF, DB	C(W, AR), V(IP)	Data	NL (R)	SFF			B	O	I, E	RA, RE	ST (E) MT	
Dandl et al. [2020]	Munich	RH	O, AB	SCI	H				Data										MPC
Dandl et al. [2021]	Virtual	RH, RS	O, IR		H			C(SR,W), V(IP)							O		RA, RE		
Dandl et al. [2022]	Munich	RH, RS	O, RB, IR		H	IND	DB,BF	C(W), V(IP)	Model	NL (R)	SFF	FleetPy	LB	IR, B	H, O	I, E	RA, RE	MT	
Dia et al. [2017]	Melbourne	RH	O, AB		H			C(SR), V(IP)	Model	NL (R)	SFF			B	H	I		MT	
Duan et al. [2020]	Manhattan	RH	O, RB		H		DB	C(SR), V(IP)	Data	NL (R)	SFF			B	O	I, E	RA, RE	ST (E)	

Citation	City	Service Type						System Model						Fleet Management					
		Sharing System	Booking Type	Electrification CI	Fleet Composition	Integration Type	Pricing	Agent Models	Demand Modeling	Transport Network	Traffic Model	TS Software	Integration Model	Dynamism	Solution Method	Vehicles Considered	Reassign, Re-optimize	Repositioning	Charging Method
Engelhardt et al. [2019]	Munich	RS	O, IR		H		DB, DeB	C(W,AR), V(IP)	Model	NL (R)	SFF			B	O	I, E	RA, RE		
Engelhardt et al. [2020]	Munich	RS	O, IR		H		DB, DeB	C(W), V(IP)	Model	NL (R)	SFF			B	H, O	I, E	RA, RE		
Engelhardt et al. [2021]	Munich	RS	O, IR		H		TB, DB	C(W), V(IP)	Model	NL (R)	SFF			B	O	I, E	RA, RE	ST (P)	
Engelhardt et al. [2022]	Manhattan	RS	O, RB, IR		H		TB, DB	C(W), V(IP)	Data	NL (R)	SFF			B	O	I, E	RA, RE	ST (E)	
Engelhardt et al. [2022]	Manhattan	RS	O, IR		H		DB	C(W), V(IP)	Data	NL (R)	SFF			IR, B	H, O	I, E	RA, RE	MT	
Erdmann et al. [2019]	Manhattan	RH	O, AB		H		TB	C(SR, AR), V(IP)	Data	GN	MM			IR, B	H, O	I	RA, RE		
Erdmann et al. [2021]	Manhattan	RH	O, IR		H		DB	C(W, AR), V(IP)	Data	NL (R)	SFF			IR, B	H, O	I, E	RA, RE	MT	
Fagnant et al. [2014]	Austin	RH	O, AB		H		TB	C(SR), V(IP)	Model	GN	EM			IR	H	I		MT	
Fagnant et al. [2016]	Austin	RH	O, AB		H		TB	C(SR), V(ItV)	Model	NL (R)	TS	MATSIM		IR	H	I, E		MT	
Fagnant et al. [2018]	Austin	RS	O, AB		H		TB	C(SR), V(ItV)	Model	NL (R)	TS	MATSIM		IR	H	I, E		MT	
Farhan et al. [2018]	Austin	RS	O, AB	SCI	H		DB	C(SR), V(IP)	Model	GN	MM			B	H	I, E			H
Fehn et al. [2019]	Munich	RH	O, AB	SCI	H		TB	C(SR), V(IP)	Model	NL (R)	SFF			IR	H	I			H
Fehn et al. [2021]	Munich	RS	O, AB		H				Data	NL (R)	TS	Aimsun							
Fielbaum et al. [2021]	Manhattan	RS	O, IR		H		TB	C(W), V(IP)	Data	NL (R)	SFF			B	O	I, E	RA, RE	ST	
Gurumurthy et al. [2018]	Orlando	RS	O, AB		H		TB	C(SR), V(IP)	Model	NL (R)	SFF			B	H	I			
Gurumurthy et al. [2019]	Austin	RS	O, AB		H	IND	BF,TB, DB,CP	C(SR), V(ItV)	Model	NL (R)	TS	MATSIM	TS	IR	H	I, E			
Hamzehi et al. [2019]	Virtual	RH	O, IR		H		DB	C(W), V(IP)	Model	NL (C)	EM				H, O				
Heilig et al. [2017]	Stuttgart	RH	O, AB		H	IND		C(W), V(ItV)	Model	NL (C)	TS	mobiiTop	TS	IR	H	I		MT	
Henao et al. [2019]	Denver	RH, RS	O, AB		H		BF, DB												
Horl et al. [2017]	Virtual	RH, RS	O, AB		H	IND	BF, DB	C(SR), V(ItV)	Model	NL (R)	TS	MATSIM	TS	IR	H	I, E	RA, RE		
Horl et al. [2019]	Zurich	RH	O, AB		H		DB	C(SR), V(ItV)	Model	NL (R)	TS	MATSIM		IR	H	I, E	RA, RE	MT	
Horl et al. [2019]	Paris	RH	O, AB		H	IND	DB	C(SR), V(ItV)	Model	NL (R)	TS	MATSIM	TS	IR	H	I, E			
Hoseb et al. [2022]	Chicago	RH	O, RB		H		TB, DB	C (SR), V(IP)	Data	CP	MM			B	H, O	I, E			
Hyland et al. [2018]	Chicago	RH	O, AB		H		DB	C(SR), V(IP)	Model	GN	MM			IR, B	H, O	I, E	RA, RE		
Hyland et al. [2020]	Manhattan	RH, RS	O, AB		H		DB, TB, DeB	C(SR), V(IP)	Data	NL (R)	SFF			B	O	I, E	RA, RE		
Iacobucci et al. [2019]	Tokyo	RH	O, AB	SCI	H		TB	C(SR), V(IP)	Model	NL (C)		Matlab		B	O	I			MPC

Citation	City	Service Type						System Model						Fleet Management					
		Sharing System	Booking Type	Electrification CI	Fleet Composition	Integration Type	Pricing	Agent Models	Demand Modeling	Transport Network	Traffic Model	TS Software	Integration Model	Dynamism	Solution Method	Vehicles Considered	Reassign, Re-optimize	Repositioning	Charging Method
Jaeger et al. [2018]	Singapore	RH	O, IR		H		TB	C(W), V(IP)	Data	NL (R)	SFF			B	O	I			
Javanshour et al. [2019]	Melbourne	RH	O, RT		H		TB	C(W), V(ItV)	Model	NL (R)	TS	Mobility Simulation		IR	H	I		ST (E)	
Jung et al. [2014]	Seoul	RH	O, IR	SCI			TB	C(W), V(IP)	Model	NL (R)	SFF			IR, B	H, O		RE		H
Jung et al. [2016]	Seoul	RS	O, AB		H		TB	C(SR), V(IP)	Model	NL (R)	SFF			B	H		RA, RE		
Kucharski et al. [2020]	Amsterdam	RS	O, AB		H		DB, DeB	C(SR), V(IP)	Data	NL (R)	SFF			B	O	I, E			
Li et al. [2019]	New York City	RH	O, RT	SCI	H		DB	C(W), V(IP)	Data	NL (C)	EM			IR	H	I		MT	H
Li et al. [2021]	New York City	RH	O, AB	SCI	H		DB	C(SR), V(IP)	Data	NL (R)	SFF			IR	H	I		MT	H,O
Liu et al. [2018]	Sioux Falls	RS	O, AB		H		TB	C(SR), V(IP)	Model	NL (C)	SFF			IR	H	I, E			
Liu et al. [2020]	Chicago	RH	O, AB		H		DB	C(W), V(IP)	Data					IR	H	I			
Loeb et al. [2018]	Austin	RH	O, AB	SCI	H		DB	C(SR), V(ItV)	Model	NL (R)	TS	MATSIM		IR	H	I			H
Lokhandwala et al. [2018]	Manhattan	RS	O, AB		H		TB	C(SR), V(IP)	Data	NL (R)	SFF			IR	H	I, E			
Ma et al. [2017]	Manhattan	RH	O, AB		NH			C(SR), V(IP)	Model	NL (C)	EM				O	I,E			
Maciejewski et al. [2018]	Berlin	RH	O, AB		H		DB			NL (R)	TS	MATSIM		IR	H	I			
Martinez et al. [2017]	Lisbon	RS	O, RB		NH	IND		C(SR,AR), V(IP)	Model	NL (R)	SFF		LB	B	H, O	I, E			
Moavenzadeh et al. [2018]	Boston	RH, RS	O, AB		NH	IND		V(ItV)	Model	NL (R)	TS		TS						
Nguyen-phuoc et al. [2023]	Singapore	RH, RS	O, AB		NH	IND	TB	C(SR), V(ItV)	Model	NL (R)	TS	SimMobility	TS	IR	H	I, E			
Pantelidis et al. [2022]	Brooklyn	RH	O, AB	SCI	H		DB	C(SR), V(IP)	Data	NL (C)	EM			IR	H	I		MT	H
Pavone et al. [2012]	Virtual	RH	O, AB		H		TB	C(SR), V(IP)	Model	NL (C)	EM			B	O	I		MT	
Pavone et al. [2015]	New York City, Singapore	RH	O, AB		H		DB	C(SR), V(IP)	Data, Model	NL (C)	MM			IR	H	I			
Pinto et al. [2020]	Chicago	RS	O, AB		H	IND	TB, DB, BF	C(SR), V(IP)	Model	NL (R)	SSF		LB	B	O	I, E			
Ruch et al. [2018]	San Francisco	RH	O, AB		H		DB		Data	NL (R)	SSF	AMoDeus (MATSIM)		B	H, O	I, E		MT	
Salazar et al. [2018]	Manhattan	RH	O, AB		H	FN	TB, CP	C(SR), V(IP)	Data	NL (R)	SSF		LB	B	O	I, E		MT	
Santi et al. [2014]	Manhattan	RS	O, AB		H		TB	C(SR), V(IP)	Data	NL (R)	SSF				O	I			
Santos et al. [2015]	São Paulo	RH, RS	O, AB		H		TB, SRS							B	H	I, E			
Shen et al. [2018]	Singapore	RS	O, AB		H	FL	DB, TB, DeB	C(SR), V(IP)	Model	NL (R)	SSF		LB	IR	H	I, E			
Syed et al. [2019]	Manhattan	RH	O, IR		H		TB	C(W), V(IP)	Data	NL (R)					H	I			
Syed et al. [2019]	Munich	RH	O, IR		H		TB	C(W), V(IP)	Data	NL (R)					H	I			

A. Overview Table for the Literature Reviewed

Citation	City	Service Type						System Model						Fleet Management					
		Sharing System	Booking Type	Electrification CI	Fleet Composition	Integration Type	Pricing	Agent Models	Demand Modeling	Transport Network	Traffic Model	TS Software	Integration Model	Dynamism	Solution Method	Vehicles Considered	Reassign, Re-optimize	Repositioning	Charging Method
Syed et al. [2019]	Manhattan	RH	O, RT		H		TB	C(W), V(IP)	Data	NL (R)	SFF			B	H	I, E	RA, RE		
Syed et al. [2021]	Manhattan	RH	O, IR		H		BF, DB	C(W), V(IP)	Data	NL (R)	SFF	FleetPy		B	O	I, E			
Syed et al. [2021]	Manhattan	RH	O, IR		H		BF, DB	C(W), V(IP)	Data	NL (R)	SFF	FleetPy		B	O	I, E		MT	
Turan et al. [2020]	Manhattan, San Francisco	RS	O, AB	SCI	H		TB, DeB, SP	C(SR), V(IP)	Data	NL (C)	AVG			B	H	I			
Wang et al. [2014]	Singapore	RH	RB, AB		H		TB	X	Model	NL (R)	SFF	Quadstone 2009		IR	H	I, E	RA		
Wang et al. [2022]	The Hague	RH	O, AB		H		BF, TB, DF	C(SR,AR), V(IP)	Model	NL (R)	TS			IR, B	H, O				
Wen et al. [2018]	A European city	RS	O, RB, IR		H	FL	BF, TB, DB, SRS	C(W)	Model	X	X	X	X	B	H	I, E		MT	
Wilkes et al. [2021]	Eggenstein Leopoldshafen	RS	O, IR		H	IND	BF, DB, DeB	C(W), V(IP)	Model	NL (R)	SFF	FleetPy mobiTopp	LB	IR, B	H, O	I, E	RA, RE		
Winter et al. [2021]	Amsterdam	RH	O, AB		H		DB	C(SR), V(IP)	Model	NL (R)	TS	MATSIM		IR	H	I		ST (E)	
Zgraggen et al. [2019]	Manhattan	X	O, AB		H	FN	TB, DB	C(SR), V(IP)	Data	NL (R)	SFF		LB	B	O	I, E		MT	
Zhang et al. [2015]	Virtual	RS	O, AB		H		TB, DB, DeB	C(SR), V(IP)	Model	GN	MM			IR	H	I, E			
Zhang et al. [2016]	Manhattan	RH	O, AB		H		DB	C(SR), V(IP)	Data	NL (C)	MM			B	O	I		MT	
Zhang et al. [2016]	Manhattan	RH	O, AB	SCI	H		TB, DB	C(SR), V(IP)	Data	NL (C)	MM			B, IR	H, O	I		MT	MPC
Zhang et al. [2017]	Atlanta	RH	O, AB		H		BF, TB	C(SR), V(IP)	Model	NL (R)	SFF			B	O	I			
Zhang et al. [2022]	Munich	RS	O, IR	SCI, PCI	H		DB	C(W), V(IP)	Data	NL (R)	SFF	FleetPy		B	O	I, E	RA, RE	MT	H
Zwick et al. [2020]	Hamburg	RS	O, IR		H		TB, DeB	C(W,AR), V(ItV)	Data	NL (R)	TS	MATSIM		IR, B	H, O	I, E		MT	
Wallar et al. [2018]	Manhattan	RS	O, IR		H		TB	C(W), V(IP)	Data	NL (R)	SFF			B	O	I, E	RA, RE	ST (E)	

Table A1.: Overview of the AMoD literature reviewed.

B. Benchmark Methods

The dissertation tested multiple repositioning methods before selecting a benchmark method for evaluation. The following presents the formulations used for testing.

B.1. Pavone's Method

PAVONE et al. [2012] developed multiple repositioning methods. Out of these, the dissertation uses the best-performing strategy called *adaptive real-time rebalancing policy* which is referred to as Pavone's method throughout the dissertation. It is used as the benchmark method due to its powerful yet simple approach. Fundamentally, its balances the surplus vehicles among individual regions as much as possible while minimizing the repositioning VKT. The dissertation adapted it according to the definition of the regional imbalances described in section 5.2.2 and uses the same mathematical notations as in section 5.2.2. First, the number of excess vehicles in each region $z \in Z$ is determined, denoted by v_z^{excess}. Since the dissertation uses two separate ways of calculating vehicle supply ($\omega_z^{sup(c)}$ and $\omega_z^{sup(v)}$), the following describes the procedure followed to calculate v_z^{excess} for both vehicle supply estimation.

The v_z^{excess} is give as:

$$v_z^{excess} = \max\left(|V_z^+| + s_z, \quad 0\right) \tag{1}$$

where for vehicle supply estimate using $\omega_z^{sup(c)}$, the s_z is calculated as:

$$\min\left(|V_z^r| + |R_z^+| - |R_z^-|, \quad 0\right) \tag{2}$$

and for vehicle supply estimate using $\omega_z^{sup(v)}$, the s_z is calculated as:

$$\min\left(|V_z^r| - |R_z^-|, \quad 0\right) \tag{3}$$

Compared to Eq. 3, Eq. 2 additionally uses the forecast of customer destinations to calculate the excess vehicles. Thus, Eq. 2 assumes that vehicles will be available in a region due to dropping off customers (R_z^+) in addition to the enroute vehicles from the last call to repositioning (V_z^r). Excess vehicles that can be repositioned are currently idle vehicles. However, in Eq. 2 and Eq. 3, if the forecast of outgoing vehicles (R_z^-) is larger than the forecast of incoming vehicles, the excess vehicles should be reduced according to Eq. 1. Contrary to PAVONE et al. [2012], a negative value for the excess vehicles is not considered as shown in Eq. 1.

Given above, Pavone's method tries to equally balance the excess vehicles among regions while minimizing the repositioning distance, given as:

$$\min_{\boldsymbol{u}} \sum_{i,j} c_{ij} u_{ij} \tag{4a}$$

$$\text{s.t. } v_i^{excess} + \sum_{j \neq i} (u_{ji} - u_{ij}) \geq v_d \qquad \forall i \in Z \tag{4b}$$

$$v_d = \frac{\sum_i v_i^{excess}}{|Z|} \tag{4c}$$

where $\boldsymbol{u} \in \mathbb{Z}_{\geq 0}^{|Z^+| \times |Z^-|}$ is the matrix of flow variables and C is the distance matrix between zone centroids. Eq. 4b guarantees that the excess vehicles are distributed equally among regions. The objective function in Eq. 4a minimizes the repositioning distance.

B.2. Wallar's Method

The second method is based on the work of WALLAR et al. [2018]. Similar to Pavone's approach, this method is also based on discrete optimization. It tries to maximize the number of customers received by a vehicle after it reaches the repositioning destination zone. However, in contrast to Pavone's approach, this method utilizes the travel times between the zone centroids.

They first calculate the rate of new customers in each region by dividing the potential number of customer origins in each region by the forecast horizon. The dissertation adopted this according to the definition of regional imbalance described in section 5.2.2. The customer rate for zone $z \in Z$ is defined as:

$$\beta_z = \max\left(0, \frac{|R_z^-| - |R_z^+|}{\Delta T_h}\right) \tag{5}$$

where R_z^+, R_z^- and ΔT_h are as described in section 5.2.2. Then the repositioning method is given by the following optimization problem:

$$\max_{\boldsymbol{u}} \left(\sum_i \sum_j (\Delta T_r - c_{ij}^t) \cdot u_{ij} \cdot \beta_j \right) \tag{6a}$$

$$\text{s.t.} \ \sum_j u_{ij} \leq |V_i^+| \qquad\qquad \forall i \in Z \tag{6b}$$

$$\sum_j \left[u_{ij} \cdot (\Delta T_r - c_{ij}^t) \right] \geq 0 \qquad\qquad \forall i \in Z \tag{6c}$$

$$\sum_i \left[u_{ij} \cdot (\Delta T_r - c_{ij}^t) \right] \leq \beta_j \cdot \rho \cdot (\Delta T_r)^2 \qquad\qquad \forall j \in Z \tag{6d}$$

where $\boldsymbol{C^t}$ is the travel time matrix between the zone centroids and $\boldsymbol{u} \in \mathbb{Z}_{\geq 0}^{|Z| \times |Z|}$ is the variables matrix for flow of vehicles. The objective function in Eq. 6a multiplies the time remaining after reaching the destination zone by the rate of customers. Thus, it maximizes the number of customers seen by the repositioned vehicle in the destination zone. Eq. 6b makes sure that number of repositioned vehicles is less than the available (idle) vehicles in the zone. Eq. 6c guarantees that the repositioning vehicles are able to reach the destination zone before the next repositioning call. Eq. 6c limits the over-saturation of vehicles in the deficiency zone. WALLAR et al. [2018] introduced the parameter ρ in Eq. 6d to control the amount of allowed over-saturation. The dissertation set the value of ρ to 1 in the simulations.

B.3. Fagnant's Method

FAGNANT and KOCKELMAN [2014] introduced multiple iterative heuristics for the repositioning of idle vehicles. The heuristics dealt with the repositioning problem by dividing the operational area into two levels of regular grids. They studied combinations of different heuristics and concluded that that the heuristic scheme (named as R1 by FAGNANT and KOCKELMAN [2014]) dealing with the coarsest level of the grid only performed the best. Thus, the dissertation only evaluated the R1 scheme.

They first calculate the *block-balance* value for each region. Similar to the above description of Pavone's and Wallar's method, the dissertation adopts the definition of block-balance according to the quantities used in the definition of imbalance weights. Thus, the block-balance for a region $z \in Z$ is given as:
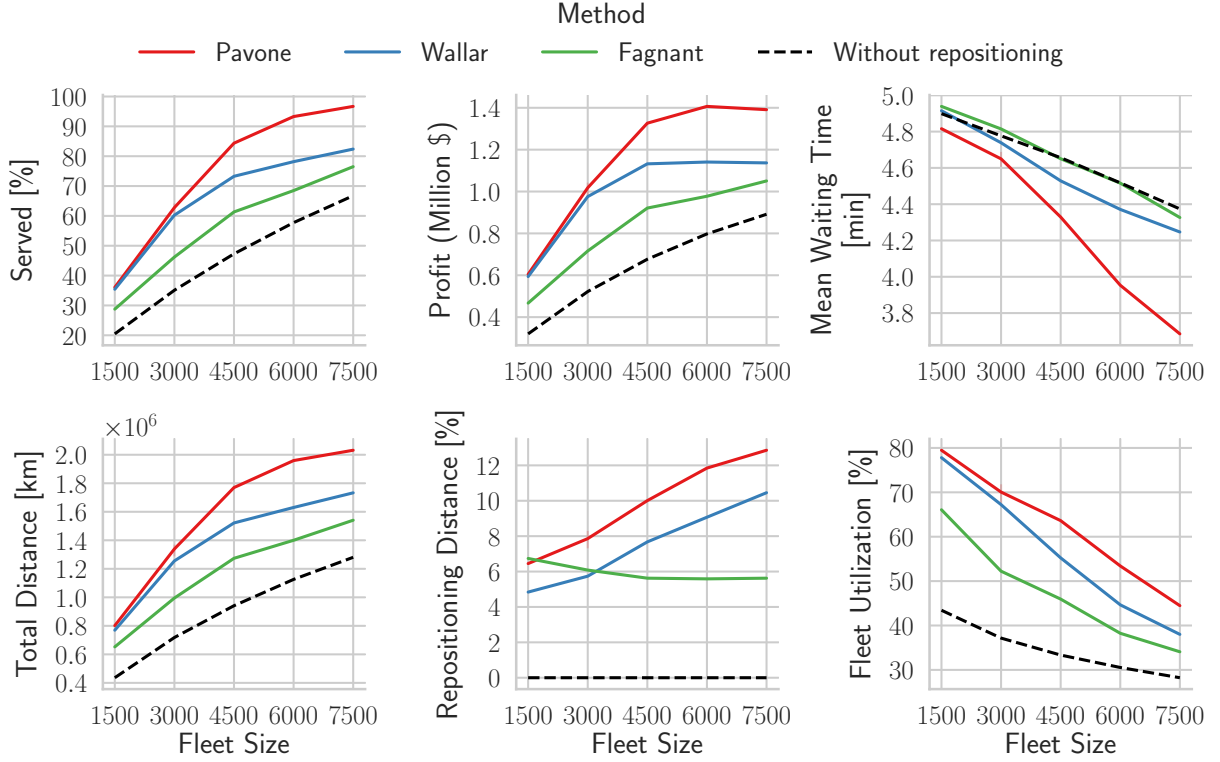
Figure B1.: Performance comparison of the benchmark methods.

$$BlockBalance_z = V_{total} \cdot \left(\frac{|V_z| + |R_z^+|}{V_{total}} - \frac{|R_z^-|}{\sum_k |R_k^-|} \right). \tag{7}$$

where $V_{total} = \sum_k (|V_k| + |R_k^+|)$.

Since the whole procedure is iterative, involving multiple steps, the description of the whole R1 method is beyond the scope of the dissertation. The only difference in our implementation is the change mentioned in Eq. 7. For details of the rest of R1 procedure, refer to [FAGNANT and KOCKELMAN, 2014].

B.4. Performance Comparison

This section compares the performance of the above benchmark methods considered in the dissertation. The simulation results presented here use the default parameters used throughout the dissertation for repositioning; refer to Table 5.1 for detail. In summary, the simulations used dynamic travel times with the repositioning period ΔT_r and the forecast horizon ΔT_h set to 30 minutes. The grid cell size Δs_{cell} of 1000 m and the maximum allowed delay ΔT_{max} of 6 minutes are used. $VCO_{enroute}$ is used for vehicle assignment. For Pavone's method, the vehicle supply is estimated using $\omega_z^{sup(c)}$. Wallar's and Fagnant's methods presented in previous sections are already adapted in an equivalent formulation to $\omega_z^{sup(c)}$.

As shown in Figure B1, all repositioning methods improve the AMoD performance; however, Pavone's approach provides the best performance, followed by Wallar's method. A major reason for higher performances is the use of optimization in both methods, while Fagnant's strategy is based

on a heuristic. For Wallar's method, maximizing the possible customers received by the repositioning vehicles does not provide the best AMoD performance. A possible reason could be the hard constraint in Eq 6c that the repositioning vehicles must be able to reach the destination before the start of the next repositioning call. This significantly limits the repositioning potential since some vehicles might be stationary in far-off regions (for example, the north part of Manhattan) and cannot reach the demand-intensive regions within ΔT_r. This is especially relevant to the simulation setup used in the dissertation due to realistic travel times. However, the supply and demand estimations can significantly affect the repositioning performance, as observed for other repositioning methods in the dissertation. Therefore, Wallar's method may perform better with different supply and demand estimations. Nevertheless, for the simulation environment used in the dissertation, Pavone's method served more customers, produced higher AMoD profit, and had lower customer waiting times. Consequently, it was selected as the benchmark method throughout the dissertation.