Fostering synergy between Public Transportation and Autonomous Mobility on Demand: the prospects of regulation

Mélanie Cortina

Thesis National Number: 2023ENTP0010

A THESIS OF ENTPE
Member of the Université de Lyon

Doctoral school No. 162
MEGA (Mechanics, Energy, Civil Engineering and Acoustics)

To obtain the graduation of
PhD in Civil Engineering

Defended on November 6th 2023 by:

Mélanie CORTINA

Fostering synergy between Public Transportation and Autonomous Mobility on Demand: the prospects of regulation

---

In front of the following examination committee:

Hervé RIVANO	Professor (INSA Lyon)	Committee chair
Eleni VLAHOGIANNI	Professor (National Technical University of Athens)	Reviewer
Oded CATS	Professor (Delft University of Technology)	Reviewer
Virginie LURKIN	Assistant Professor (Université de Lausanne)	Examiner
Lampros YFANTIS	Engineer (Aimsun)	Examiner
Ludovic LECLERCQ	Professor (ENTPE)	Examiner
Nicolas CHIABAUT	ITPE engineer HDR (Département de la Haute Savoie)	Supervisor

Thesis prepared at the LICIT-ECO7 laboratory

Numéro national de thèse (NNT): 2023ENTP0010

THÈSE DE DOCTORAT DE L'ENTPE
Membre de l' Université de Lyon

École doctorale No. 162
MEGA (Mécanique, Énergétique, Génie civil et Acoustique)

Spécialité / discipline de doctorat
Génie civil

Soutenue publiquement le 06/11/2023, par:

Mélanie CORTINA

Perpectives de régulation pour encourager les synergies entre les transports en commun et la mobilité autonome à la demande

Devant le jury composé de:

Hervé RIVANO	Professeur (INSA Lyon)	Président
Eleni VLAHOGIANNI	Professeure (National Technical University of Athens)	Rapporteure
Oded CATS	Professeur (Delft University of Technology)	Rapporteur
Virginie LURKIN	Professeure assistante (Université de Lausanne)	Examinatrice
Lampros YFANTIS	Ingénieur (Aimsun)	Examinateur
Ludovic LECLERCQ	Professeur (ENTPE)	Examinateur
Nicolas CHIABAUT	Ingénieur ITPE HDR (Département de la Haute Savoie)	Directeur de thèse

Thèse préparée au laboratoire LICIT-ECO7

"One of the axioms of complexity is
the impossibility, even in theory,
of omniscience."

Edgar Morin

Abstract

Shared mobility is part of the answer to today's environmental challenges but should face some limitations, such as supply-demand imbalances, stochasticity in supply, and unaffordable rides. Autonomous vehicles could help overcome these limitations. But the promises of Autonomous Mobility on Demand (AMoD) regarding the environmental challenges are not to be taken for granted. To prevent a too-high induced demand and an increase of the vehicle kilometers traveled (VKT), integrating AMoD and Public Transportation (PT) is needed. However, private companies are the most likely to be the first equipped with large fleets of AVs capable of meeting urban demand. The cooperation of AMoD and PT is not ensured. Hence, regulation measures are needed to achieve the benefits of intermodal AMoD.

The main issue addressed in this thesis is how to regulate AMoD to foster cooperation with PT? This question is tackled in two different study cases, a transportation corridor and a large urban area. The work has three main objectives. First, it accounts for the benefits of a multimodal system based on the cooperation between PT and AMoD regarding efficiency, sustainability, and equity. Second, it aims at understanding the circumstances of cooperation/competition between PT and AMoD. The idea is to identify under which conditions AMoD cooperates or competes with PT and describe the associated mobility patterns. Third, we propose optimized means to realize the benefits of intermodal AMoD. It consists in optimizing the regulation strategies chosen for both case studies.

Contributions of the work are threefold. To begin with, we test and optimize new regulation policies that maximize the social and environmental benefits of intermodal AMoD. In the corridor case study, we investigate the joint PT design and AMoD service disaggregation into fleets operating on geofenced coverage zones. In the urban area case study, we explore several combinations of monetary measures affecting both travelers and autonomous vehicles.

An ideal approach to treat the research question would account for a flexible intermodality, the network loading dynamics, a detailed AMoD model, and the coupling between supply and demand while remaining amenable to the optimization of the regulation. We propose new and refine existing modeling bricks necessary to build such an approach. In the corridor case study, we introduce a simple dynamic model incorporating time-dependent mode and route choice subject to user equilibrium constraints and amenable to optimization. In the urban area case study, we propose and analyze a batch-matching on horizon operational policy for AMoD that integrates well into an agent-based simulation approach.

Finally, we integrate the different modeling bricks together to get closer to the ideal approach for optimizing AMoD-specific regulation measures.

RÉSUMÉ

La mobilité partagée est une réponse possible aux enjeux environnementaux soulevés dans le secteur des transports urbains. Pour qu'elle contribue réellement à la diminution de l'empreinte carbone du système sans mettre en péril son efficacité, elle doit faire face à certaines barrières telles que les déséquilibres inhérents aux motifs de mobilité urbains, la compléxité de la gestion optimale de flotte en présence d'une demande et d'une offre incertaines, ou encore le coût important des trajets à la demande. Les avancées technologiques récentes en matière de conduite autonome peuvent nous laisser envisager des services de mobilité à la demande basés sur les véhicules autonomes. Si la mobilité autonome à la demande (MAD) pourrait permettre de surmonter les limitations citées plus haut, d'autres risques apparaissent, notamment l'augmentation du nombre de kilomètres parcourus en raison des relocalisations à vide, et la génération d'une demande induite importante. En fait, la MAD n'est pertinente que lorsqu'elle fonctionne en synergie avec les transports en commun (TC), plus efficaces pour dégager des économies d'échelle. Dans le cadre d'une MAD privée, à l'image des services à la demande existant tels q'Uber, la coopération avec les TC n'est pas garantie. En effet, les prises de décision opérationnelles des usagers (choix de mode et d'itinéraire) et du gestionnaire de MAD (dispatch et affectation des véhicules) sont généralement motivés par des objectifs qui leur sont propres et non par l'aspect environemental. Ainsi, la question de la régulation de la MAD se pose.

Cette thèse est donc motivée par la question suivante: comment réguler de manière optimale la MAD pour favoriser sa coopération avec les TC? Cette problématique est étudiée dans deux cas d'études distincts, un couloir de mobilité et une aire métropolitaine, à l'aide de deux approches différentes. Sur chacun de ces cas, le travail répond à trois objectifs. Premièrement, il s'agit de rendre compte des bénéfices de la complétion des TC par la MAD en termes d'efficacité du système (temps de parcours, coût de trajet, individuels et cumulés), d'indicateurs environnementaux (part de la mobilité collective dans les déplacements, émissions carbone), et d'équité. Deuxièmement, nous cherchons à mettre en évidence les circomstances au cours desquelles MAD et TC coopèrent ou sont en compétition et de décrire les motifs de mobilité correspondant. Enfin, il s'agit de proposer des politiques de régulation optimales, capables d'exploiter le plein potentiel de la MAD intermodale.

Ce travail de thèse présente trois types de contribution. Le premier consiste à tester des politiques de régulation dédiées à la MAD relativement peu étudiées dans la litérature, notamment avec une approche en optimisation et avec les fonctions objectifs qui nous occupent. Dans le cas du couloir de mobilité, nous étudions la désaggrégation de la MAD en plusieurs flottes opérant dans des zones de couverture bien distinctes. Dans le cas de l'aire métropolitaine, nous explorons plusieurs combinaisons de taxes et subventions adressées aux voyageurs et aux véhicules autonomes.

Le second consiste à proposer de nouvelles approches et à retravailler certaines briques de modélisation permettant d'étudier les intéractions entre la MAD et les TC. Nous pro-

posons un modèle simple et efficace pour le calcul de l'équilibre usager dynamique dans le couloir de mobilité. Dans le cas de l'aire métropolitaine, nous proposons et étudions en détails plusieurs stratégies opérationelles de dispatch des véhicules autonomes rendant compte de l'objectif de l'opérateur de la MAD, tourné vers la maximisation de son profit.

Enfin, la dernière contribution est l'intégration des différentes briques de modélisation proposées dans une approche en simulation qui se rapproche d'une méthodologie exhaustive, prenant à la fois en compte une intermodalité flexible, la dynamique du trafic, les contraintes opérationnelles de la MAD et le couplage entre la demande et l'offre tout en restant compatible avec l'optimisation de la régulation.

ACKNOWLEDGEMENTS

First, I would like to thank Nicolas Chiabaut, without whom I could not have discovered the world of scientific research. You offered me the opportunity to work on an exciting subject with the highest freedom possible. From the few months we could work together, I'll always remember your "not important, not sure, manageable" mantra that helped me on several occasions during these three years. I regret the short time we had to collaborate in proper conditions between the sanitary crisis and your departure from the laboratory. Thank you for your help kicking this project off and your support till then.

My second thoughts go to Prof. Ludovic Leclercq, who agreed to take over from day-to-day mentorship. I want to thank you for your support and encouragement. On the scientific front, I highly appreciated working with you and benefiting from your expertise and sound advice. Your brilliant ability to quickly and clearly understand a problem briefly and fuzzy explained to you has been a real aid till the end. On the human front, I warmly thank you for the time you took to understand my operating mode, to cheer me up when doubt was a brake, and for your due care and honesty.

I address my deep and warm acknowledgments to Prof. Eleni Vlahogianni from the National Technological University of Athens and Prof. Oded Cats from the Delft University of Technology for agreeing to review my manuscript thoroughly. Their insights and perspectives have been valuable in examining my research work. I also thank the other panel members for their careful examination: Prof. Virginie Lurkin from Lausanne, University, Mr. Lampros Yfantis from Aimsun, and Prof. Hervé Rivano from INSA Lyon who agreed to chair the committee.

During the last three years, I have learned much thanks to my colleagues at the LICIT-ECO7 laboratory. This big family has been welcoming, and I was proud to be part of this team of enthusiasts. I want to extend my warmest greetings to all of them, those with whom I shared a discussion, a sport session, a beer, delicacies, hints on MnMS, trips for conferences, Molkky games, complaints about ENTPE firewall, writting tips, etc. : Cyril, Anna, Manon, Carlos, Ruiwei, Mahendra, Jean, Mathis, Alexandre, Mina, Florian, Cécile, Delphine, Christine, Silvia, Michele, Thibault, Christophe, Flore, Nicolas in ENTPE, and Andres, Louis, Pierre, Fayçal, Matthieu, Tejas, Serge, Bruno, Ali, Pierre-Antoine, Angelo, Loïc, Éléonore, Hugues, Jiajie, Cécile, Maryia in Gustave Eiffel University. I especially thank Sonia and Anne-Christine for their respective support, kindness, and availability.

I also had the chance to exchange with the École Urbaine de Lyon colleagues and participate in communication, knowledge diffusion, and awareness-raising activities that appeared to me so far from scientific research at that time but were definitively essential in the Anthropocene. I had much fun creating and animating radio programs, and I learned so much working with this multidisciplinary team: Jérémy, Lucas, Sofia, Rulan, Pénélope,

Adrien, Jindra, Yann, Jade, Félix, Thomas Bo., Thomas Ba. and Robert.

I want to express my last thanks to my family and friends. Greetings to Fanny for her constant support, Guillaume for believing in me whatever the circumstances, Nicolas for accepting being my Ph.D. buddy, listening to my countless worries and to my joys, making me want to do more efforts everyday while not forget to smile, Laurence and Thierry for teaching me that giving up is never the good option.

This thesis has been sponsored by École Urbaine de Lyon and ENTPE.

Contents

List of Figures

List of Tables

List of abbreviations

AMoD Autonomous Mobility on Demand.

AV Autonomous Vehicle.

BO Bayesian Optimization.

CBD Center of the Business District.

DO Drop-Off.

DUE Dynamic User Equilibrium.

FIFO First-In-First-Out.

GA Genetic algorithm.

HV Hypervolume.

MAE Mean Absolute Error.

MAES Model-assisted Evolutionary Strategy.

MARL Multi-Agent Reinforcement Learning.

MnMS Multimodal network Modeling and Simulation.

MOBO Multi-Objective Bayesian Optimization.

MoD Mobility on Demand.

MOEA Multi-Objective Evolutionary Algorithm.

MRT Massive Rapid Transit.

MSA Method of Successive Averages.

NDS Number of Non-Dominated Solution.

OzDz Origin zone - Destination zone.

PC Personal Car.

PF Pareto Front.

PS Pareto Set.

PT Public Transportation.

PU Pick-Up.

SAV Shared Autonomous Vehicles Services.

TA Transportation Authority.

TAZ Traffic Analysis Zone.

TDT Total Distance Traveled.

TNC Transportation Network Company.

UE User Equilibrium.

VKT Vehicle Kilometers Traveled.

Part I

Fostering synergy between Public Transportation and Autonomous Mobility on Demand: why and how?

1. A GENERAL INTRODUCTION

1.1 Context

1.1.1 The relevance of Autonomous Mobility on Demand to deal with the environmental challenges in urban mobility

Cities worldwide are still growing. By 2050, 68% of the global population is expected to be urban (United Nations, 2019, Melchiorri *et al.*, 2018). As a result, traveling demand increases, and mobility patterns become more complex there. The personal individual car model cannot handle such a level of demand. As a result, traffic conditions are deteriorating. Congestion causes severe economic and environmental damage. Extra travel time and fuel consumption account for 180 billion dollars in the US in 2017 (Schrank *et al.*, 2019), and 22.5 billion dollars in France in 2013 (CEBR, 2014). At the local level, fine-particle emissions and noise pollution adversely affect citizens' daily lives and health (Titos *et al.*, 2015). At the global level, extra fuel consumption and carbon emissions play a significant role in global warming (Nicolas *et al.*, 2012, André & Vieira da Rocha, 2020). US and Western European cities generate on average 4405 kg and 1269 kg of CO2 per person per year for passenger transport respectively (Kenworthy, 2003).

These environmental challenges call for planning sustainable transport and mobility within and beyond cities. The development of transit, the incentives for vehicle and fuel switching, and the promotion of active travel and collective transport are all policies to make urban mobility greener. Collective mobility, also called shared mobility, is defined as the shared use of a vehicle, motorized or not. It is particularly promising to address the gaps in public transportation (PT), discourage private car ownership, and increase vehicle utilization rates.

The concept has declined into several systems deployed in cities (Figure 1.1a). The wide range of modes belonging to shared mobility splits into two categories depending on what is being shared (Shaheen & Chan, 2016). A vehicle is shared sequentially in carsharing, scooter sharing, and bikesharing. Ridesharing and on-demand ride services enable the share of a passenger ride. There has been explosive growth in on-demand ride services during the last decade, thanks to advances in mobile technology. Compared to traditional ridesharing,

such as carpooling [1], passengers request rides through a mobile application. Ridesourcing, ridesplitting, and e-ride-hailing are three variants of on-demand ride services. In ridesourcing, a Transportation Network Company (TNC) provides an application that matches drivers with passengers. Uber, Lyft, and Didi are examples of this type of service. In ridesplitting, a ride and its associated fares are split between two or more passengers. UberPOOL proposes this kind of service. E-ride-hailing refers to passengers hailing a taxi with a mobile application for an individual ride. Ridesplitting belongs to the panel of microtransit [2] solutions.

Shared mobility systems, however, present significant limitations. For one-way carsharing, unevenly distributed origins and destinations cause spatial imbalances in the level of service. In ridesourcing and ridesplitting, the control of supply by the TNC is complex. The availability of drivers on a certain period of the day is difficult to predict (Nourinejad & Ramezani, 2020, Sun *et al.*, 2019, Zha *et al.*, 2018) and the dispatching orders are not necessarily followed by drivers, resulting in a sub-optimal distribution of supply over the demand. Moreover, rides are generally unaffordable for most travelers.

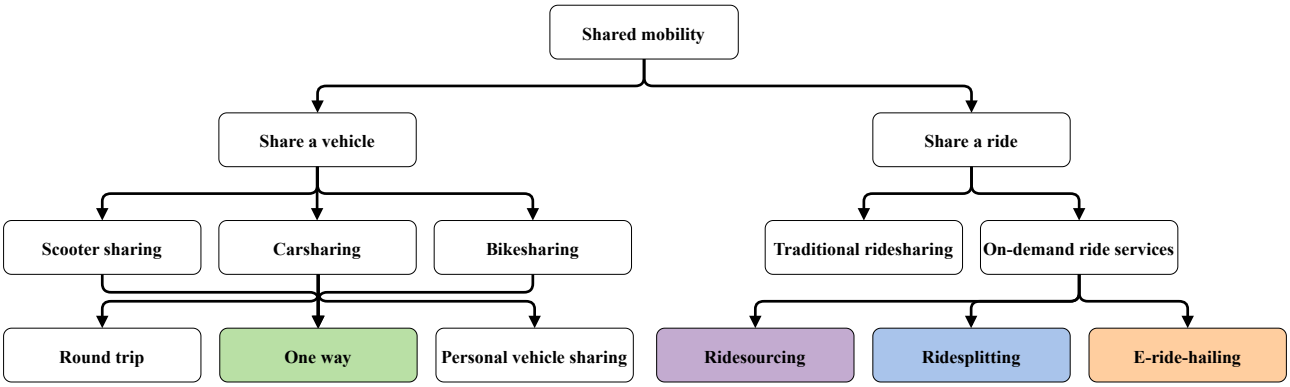
Technological advances in autonomous driving allow us to envisage the next generation of shared mobility services, overcoming these limitations. Self-rebalancing, perfect control of the fleet, and cost savings (Becker *et al.*, 2020) are possible in Shared Autonomous Vehicles Services (SAV) (Narayanan *et al.*, 2020). Consequently, car manufacturers, digital corporations, TNCs, and states invest billions of dollars yearly to develop the technology (Kerry & Karsten, 2017).

If one imagines ridesourcing systems using autonomous vehicles (AVs) exploited by third-party owners, the competition between selfish AVs will probably bring fewer benefits than cooperation (Hryhoryeva & Leclercq, 2023). When a unique company owns and manages the AVs, we speak of Autonomous Mobility-on-Demand (AMoD). The concept of AMoD can be further categorized based on serving points definition (given pickup (PU) and dropoff (DO) points, or any location PU and DO) and sharing system (only one customer in an AV at a time, or several customers share the AV at the same time). Figure 1.1b describes the usually studied AMoD systems, namely, car sharing with AVs (Pavone, 2015, Gurumurthy & Kockelman, 2022, Javanshour *et al.*, 2022), autonomous ride-hailing (Yi & Smart, 2021, Al-Kanj *et al.*, 2020), and AMoD with ride-sharing (Alonso-Mora *et al.*, 2017b, Fagnant & Kockelman, 2018, Khemiri *et al.*, 2022).

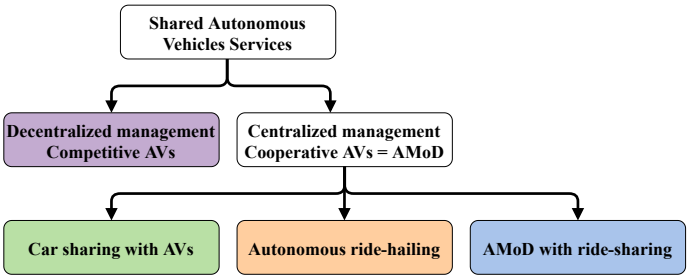
1.1.2 The need for regulating AMoD systems

Compared to carsharing and bikesharing, cities and transportation authorities' enthusiasm for AMoD is more nuanced and depends on the territory. In the US and China, Uber, DiDi, Waymo, and Baidu have launched autonomous ride-hailing pilots (Dowling & McGuirk, 2022). AVs already serve customers in some cities' streets (Figure 1.2a). In Europe, many experiments are ongoing to test the relevance of AMoD for complementing existing public transit systems or supplying sparsely populated areas. In France, the ENA project (Université Gustave Eiffel, 2019), sponsored by the French government, deployed several au-

[1] Shared rides among drivers and passengers with a compatible origin, destination, and departure time
[2] Multi-passengers transportation system operated either with fixed, flexible time tables or on-demand



(a) *Categories of shared mobility, adapted from Shaheen & Chan (2016). Autonomous vehicles are relevant for one-way carsharing, ridesourcing, ridesplitting, and e-ride-hailing systems.*



(b) *Types of SAV. We speak of "car sharing with AVs" when travelers can board and alight AVs at given pickup (PU) and dropoff (DO) points. AVs rebalance themselves between these locations. It is the application of one-way carsharing with AVs. We speak of "autonomous ride-hailing" when a private company deploys a fleet of small-sized AVs to pick up and drop off requests individually at any time and location. It is the application of e-ride-hailing with AVs. When two or more travelers can share an AV simultaneously, we use the name "AMoD with ride-sharing". It is the application of ridesplitting.*

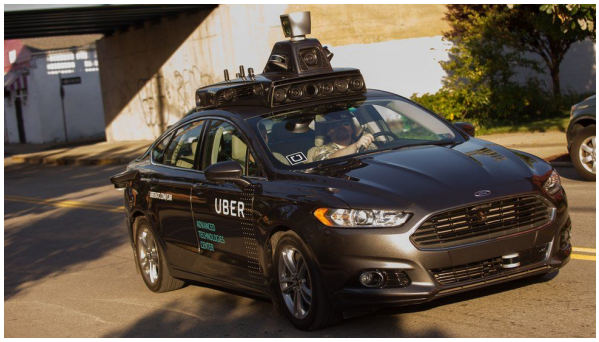
FIGURE 1.1 – *Categories of shared mobility and its application with autonomous vehicles.*

tonomous shuttle services for the last mile in connection with transit lines (Figure 1.2b).

The reason is that AMoD's promises to deal with the urban and environmental challenges are not to be taken for granted. Milakis *et al.* (2017) classify the potential impacts of AVs into three categories. First-order impacts are related to traffic, travel cost, and travel choices and should be globally positive. However, second-order impacts, related to vehicle ownership, vehicle sharing, land use, and transport infrastructures, and third-order impacts, related to energy consumption, air pollution, safety, social equity, economy, and public health, are expected to be negative or highly uncertain. Fagnant & Kockelman (2015) highlight similar watch-points. Several of them are relevant to the deployment of AMoD in cities.

In terms of safety, if autonomous driving can help reduce the number of crashes, designing a system that can perform in every possible situation is challenging. Research is still active on this topic (Riedmaier *et al.*, 2020). Moreover, to achieve safety and traffic efficiency gains, a minimum penetration rate of AVs is required and depends on the network type (Gueriau & Dusparic, 2020).

Regarding travel behavior, facilitating personal independence and mobility with AMoD could induce more demand. Travelers may be willing to commute longer distances, increas-



(a) *Uber AV in San Francisco. Credit: Getty Images.*

(b) *Autonomous shuttle of the ENA project. Credit: Sophie Jeannin.*

FIGURE 1.2 – *Deployed autonomous vehicles.*

ing urban sprawl. Other substantial changes in land use are awaited as the need for PU/DO spots in city centers will replace the need for parking (Stein, 2021).

Regarding health, a risk of negative impact exists due to a potential decrease in walking distances .

In terms of sustainability, if one shared AV could replace from 1.18 (Lang *et al.*, 2018) to 10 (Fagnant & Kockelman, 2014) conventional cars, according to the literature, the majority of studies find that the Vehicle Kilometers Traveled (VKT) will increase with the introduction of car sharing with AVs, and autonomous ride-hailing (Dia & Javanshour, 2017, Moreno *et al.*, 2018, Masoud & Jayakrishnan, 2017). Only the AMoD with ridesharing application presents encouraging results with a reduction of VKT up to 55% (Lokhandwala & Cai, 2018). Under greater demand and VKT, congestion and emissions may worsen.

A possible solution to this issue of sustainability is to integrate AMoD and PT, (Fraedrich *et al.*, 2019), i.e., taking advantage of the economies of scale achieved by mass transit while solving the first and last-mile problems with AMoD. Salazar *et al.* (2018) call such a system intermodal AMoD. Several simulation studies have shown the benefits of intermodal AMoD. In Shen *et al.* (2018), congestion and out-of-vehicle time (waiting, walking time) reduce when an AMoD with ridesharing system replaces underused feeder buses. In Zhou *et al.* (2019), a car sharing with AVs program improves the travel time of inbound commuters in the morning, outbound commuters in the evening, and elderly residents during the day. Basu *et al.* (2018) show that mass transit is irreplaceable, despite the efficiency of AMoD, to maintain a sustainable transportation system with acceptable levels of service by comparing three scenarios (no AMoD, no transit, AMoD and transit).

The example of ridesourcing has shown both complementarity and competition with PT, depending on the territory, time horizon, and data sources. Table 1.1 gathers studies that exploit travel surveys, proxies and transit ridership data, or detailed ridesourcing trips and transit passengers count data to derive correlations. Other works rely on models to estimate the potential of ride-sourcing to complement or compete with transit (Narayan *et al.*, 2019). Most studies agree on a decrease in bus ridership following TNC's entry on the market while the impact on light rail and subway ridership is variable.

Reference	Data	Spatial focus	Time horizon	Major findings
Rayle *et al.* (2016)	Travel survey	San Francisco	Spring 2014	• > 50% of ridesourcing trips replace modes other than taxi: transit, car
Dong (2020)	Travel survey	Philadelphia	Winter 2019	• 1/4 replaced transit with ridesourcing in their last trips
Nelson & Sadowsky (2019)	Proxies (date of 1st and 2nd TNC entry), transit ridership data	Several US metrop. areas	Up to 48 months after 1st and 2nd TNC entry	• Increase of transit ridership after 1st entry • Decrease of transit ridership below the original level after 2nd entry
Hall *et al.* (2018)	Proxies (date of TNC entry/exit), transit ridership data	Several US metrop. areas	2012 to 2016	• +5% transit ridership on average • Heterogeneity depending on the area and transit agency sizes
Erhardt *et al.* (2022a)	Proxies (date of TNC entry), transit ridership data	Several US metrop. areas	2012 to 2018	• -1.9% to -3.4% bus ridership per year after TNC entry • Little effect on light rail in large metrop. areas, -10% in medium ones
Erhardt *et al.* (2022b, 2019)	TNCs trips detailed data, passengers count in transit	San Francisco	2010 to 1015	• -10% transit ridership, no significant impact on light rail ridership • TNCs contribute to congestion growth in San Francisco

TABLE 1.1 – *Review of some findings on ridesourcing effect on transit ridership.*

Rayle *et al.* (2016) point out that TNC rides replace taxi trips with shorter and more reliable waiting times. Hall *et al.* (2018) argue that PT is cheaper enough that Uber's role in adding flexibility to the system is more important than its ability to substitute for transit. Nelson & Sadowsky (2019) find that PT use tends to increase immediately following the entry of Uber in the market but that the presence of two TNCs lowers prices and makes the PT ridership level fall below the initial level.

Since AMoD should ensure better management of the fleet and cheaper rides, the risk for a modal shift from PT to AV is high. Gurumurthy *et al.* (2020) confirm the risk by testing the sensitivity of AV fares on the complementariness-substitute relation between AMoD and PT. Hörl *et al.* (2021) simulate the city of Zurich with an unregulated autonomous ride-hailing system and derive states in which service cost, waiting time, and demand are in equilibrium. They conclude that the service competes with transit and active modes regardless of the fleet size tested.

A public intermodal AMoD system would be optimal in terms of control of the urban form, equity in access to mobility, coordination with land use planning, and balancing the allocation of the urban space between modes (Saujot *et al.*, 2018). However, private companies are the most likely to be the first equipped with substantial fleets of AVs capable of meeting urban demand. Then, public powers should anticipate and wonder how to take advantage of private AMoD. Regulation measures are needed to achieve the benefits of intermodal AMoD with private AMoD operators.

1.2 Thesis objectives and contributions

The main research question addressed by this thesis is: *How to regulate AMoD to foster synergy with PT?*

We answer this question in two case studies with two different modeling approaches. For each of them, three objectives are addressed:

- Account for the benefits of a multimodal system based on cooperation between PT and AMoD. We compare a baseline scenario with no AMoD and a scenario with intermodal regulated AMoD regarding system efficiency, sustainability, and equity.
- Understand the circumstances of cooperation/competition between PT and AMoD. The idea is to identify under which conditions AMoD cooperates or competes with PT and describe the associated mobility patterns.
- Propose optimized means to realize the benefits of intermodal AMoD. It consists in optimizing the regulation strategies chosen for both case studies.

The contributions of this thesis are threefold. They respectively relate to AMoD-specific regulation policies building (**C1**), the development or refinement of modeling approaches, bricks, and methodologies (**C2**), and the integration of modeling bricks together (**C3**). The following subsections provide brief state-of-the-art regarding regulation policies for AMoD and modeling approaches to treat the question of PT-AMoD interactions. They present our contributions to each of these topics.

1.2.1 AMoD-specific regulation policies (C1)

1.2.1.1 State-of-the-art

AMoD systems are hybrid. They are more flexible than PT but less than personal cars (PCs). They are more collective than PCs but less than PT. They should be better than taxis and Mobility on Demand (MoD) systems in answering unbalanced demand and taking advantage of shared rides. Regulations usually applied to PCs, taxis, and MoD systems are not necessarily relevant to AMoD. Commission cap (Vignon *et al.*, 2021), driver minimum wage (Li *et al.*, 2019), cordon toll, and congestion pricing (Li *et al.*, 2021) have been investigated in the context of human-driven vehicles. However, the study of AMoD-specific regulation measures is relatively recent. Hence, there is a need to design AMoD-specific regulations and quantify their implications.

Regulations can be categorized into push and pull measures depending on whether they reward behaviors consistent with the policy objective or punish behaviors conflicting with it. Push and pull measures, respectively, decrease and increase the utility of one or several modes. Regulations can also be categorized depending on whether they impact the system through monetary levers or by constraining the design and operation of the supply. Several types of regulation could help maximize the benefits of intermodal AMoD. Figure 1.3 presents examples of push and pull regulations that can apply to both MoD and AMoD systems.

	PUSH Disadvantage personal car, door-to-door individual AV rides	**PULL** Favors PT, intermodal AV-PT trips, active modes
MONETARY	- Per-trip tax - PU/DO tax - Imposed ridesplitting fare ratio - Congestion pricing (- Parking price) (- Cordon toll)	- Subsidize PT trip - Subsidize intermodal AV-PT trip
DESIGN OPERATION	- Limit flexibility on PU/DO location - Limit coverage zone/accessible links - Cap the number of vehicles - Imposed priority for ridesharing matches	- Minimum service level requirement - Increase PT budget (frequencies) - Improve the PT network design - High occupancy vehicles lane

FIGURE 1.3 – *Examples of regulations studied in the context of MoD and AMoD.*

Some monetary push measures apply to on-demand services only. Li *et al.* (2019) and Zhang & Nie (2019) investigate the effect of a per-trip tax in ridesourcing. The former finds that taxing trips increases fares and decreases drivers' wages and TNC revenue. The latter finds that per-trip tax can only increase social welfare under a minimum wage policy for drivers. Ke *et al.* (2021) focus on ridesourcing and shows that an optimal trip fare to maximize social welfare equals the marginal cost of using a vehicle to serve a passenger. The

tax applied to (A)MoD trips can also depend on the locations of PU and DO points. In Räth *et al.* (2021), the more accessible a PU/DO point is with PT, the higher the PU/DO tax is. Applying this scheme leads to a slight mode shift from AMoD to PC, PT, and active modes. In the context of ridesplitting, Zhu *et al.* (2020) show that increasing the ridesplitting fare ratio and prioritizing the matching of shared rides over single rides helps maintain the PT demand.

Other monetary measures apply to personal cars (PCs) and (A)MoD. A widespread regulation is congestion pricing. It charges vehicles for the delays or congestion they cause. Simoni *et al.* (2019) analyze the effects of different congestion pricing strategies in a scenario with autonomous ride-hailing. They test advanced congestion pricing schemes (global and link-based dynamic tolls) and traditional ones (flat tolls on the most congested links and distance-based tolls on all links). They show that the advanced schemes are not better than the traditional ones in affecting demand and traffic but bring higher social welfare. Salazar *et al.* (2018, 2020) propose an optimal link-based road toll that maximizes social welfare in an intermodal AMoD system by reducing mean travel times, mean travel costs, and emissions. Dandl *et al.* (2021) optimize a dynamic road toll scale factor, where the toll is distance-based, and its coefficient increases linearly with vehicle density within the city.

Finally, some push measures only apply to PCs and complement the effects of other regulations. For example, Dandl *et al.* (2021) consider a combination of regulations, including parking fees for PCs. The same goes for Räth *et al.* (2021) who consider a cordon toll for PCs.

A certain number of push measures constrain the design and operation of AMoD. In Gurumurthy *et al.* (2020), the role of autonomous ride-hailing sticks to first-mile and last-mile trips because the flexibility in PU and DO locations is limited. It increases PT coverage, lowers average access and egress walking distances, and shifts demand away from park-n-rides and long walk trips. To preserve PT mode share, Räth *et al.* (2021) restrict the AMoD service area by excluding the zones where PT accessibility during rush hours is above a certain threshold. Gurumurthy *et al.* (2021) also limit the flexibility of an AMoD by geofencing its service area. Comparing a few scenarios shows that appropriate boundaries for the service area can decrease empty VKT and response times. Li *et al.* (2019) limit the level of service of a ridesourcing system by capping the number of vehicles in the fleet. They highlight the risk for TNC to reap all the benefits of limiting supply.

Regarding pull measures, subsidizing PT and intermodal PT-AV trips are intuitive measures. In the context of ridesourcing, Reck & Axhausen (2019) argue that long first and last-mile trips and high-income households are more likely to benefit from a fixed subsidy provided to travelers for first and last-mile AV trips in connection with specific PT stations. Zhu *et al.* (2021) show that TNC's profit in areas with high PT accessibility can decrease with a fixed subsidy attributed to passengers choosing ridesourcing for the first mile. In Salazar *et al.* (2018, 2020), link-based PT fares are jointly optimized with road tolls to achieve the benefits of intermodal AMoD and reach the social optimum.

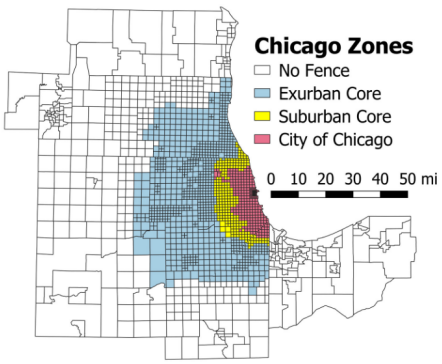
Few studies deal with pull measures on the design and operation of supply. Gao & Li (2023) study the effect of a minimum service level requirement imposed on AMoD. With a game-theoric model applied to San Francisco County, they show that such a directive improves spatial equity and social inequity. Dandl *et al.* (2021) propose increasing the PT budget, which means raising the PT lines frequencies, decreasing their crowding, and im-

proving travelers' willingness to transfer. Mo *et al.* (2021) also consider PT lines frequency as a decision variable to design efficient intermodal AMoD systems. Pinto *et al.* (2020), Kumar & Khani (2022), Basciftci & Van Hentenryck (2023) go further by including the decision of which PT line to operate. The PT network design becomes a lever to foster synergy between PT and (A)MoD.

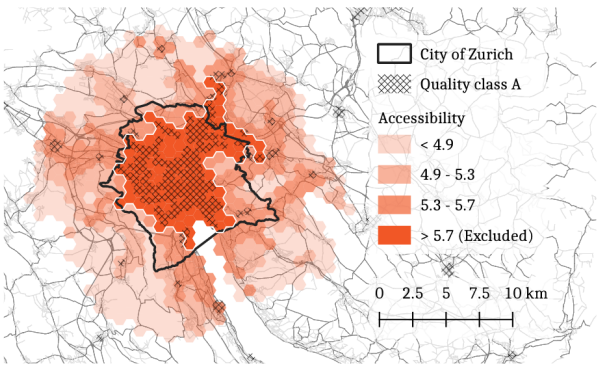
1.2.1.2 Contributions C1

In this thesis, we test and optimize policies that exploit the possible synergy between PT and AMoD to maximize the environmental benefits of AMoD deployment in urban and suburban areas. They take into account the diversity of stakeholders and their respective objectives. Each regulation proposed befits the study case in which it is investigated.

In part II of the thesis, **we tackle the morning commute in a transportation corridor where limiting AMoD fleets' coverage zones is a particularly appropriate push measure**. Gurumurthy *et al.* (2021) tests only four geofencing scenarios, including the city of Chicago, the suburban core, the exurban core, and all Chicago regions' Traffic Analysis Zones (TAZs) (Figure 1.4a). Räth *et al.* (2021) defines only one accessibility threshold above which a hexagonal cell is excluded from AMoD's service zone(Figure 1.4b). In our corridor case, the regulator looks for the best way to geofence AMoD. We optimize the number of fleets operating along the corridor and the boundaries within which each fleet operates. On top of regulating AMoD, we consider a pull measure regarding PT design. The PT design decision variables are the number of stations and their locations on a Massive Rapid Transit (MRT) line operating in the corridor. **We combine the AMoD regulation and the PT design problems by formulating a joint PT-AMoD design problem in chapter 3. To the best of our knowledge, this combination of decision variables has not been investigated yet in the literature.** Such a joint regulation and design strategy may achieve significant gains in transportation corridors. Moreover, **we consider several possible priority objectives for the regulator and question a scenario accounting for the divergence between the regulator's objective and the private AMoD operator's one**.



(a) *Spatial extent of AMoD coverage zones (stacked) in the Chicago region (from Gurumurthy et al. (2021)).*



(b) *Coverage zone of AMoD the in Zurich region excludes hex cells with an accessibility score above 5.7 (from Räth et al. (2021))*

FIGURE 1.4 – *Examples of limiting AMoD coverage zone in literature.*

In part III of the thesis, **we still tackle the morning commute but in a large urban area**.

The coverage zone geometries would be more complex to define in such territory. Moreover, several types of PT operate in the city, and refining the entire PT network design is tremendously challenging and unrealistic. Consequently, another regulation strategy, which is more suited to the case, is considered. We test new combinations of monetary push and pull measures. Concretely, we introduce five flat prices: a tax for travelers using AMoD for a door-to-door ride, a subsidy for travelers using AMoD for an intermodal AV ride, a tax on PC trips, a tax for AVs serving a door-to-door ride, and a subsidy for AVs serving an intermodal ride. Our pricing scheme goal is to incentivize travelers to use AVs in connection with PT instead of PCs and disincentivize using AVs when PT is available. It also aims to incentivize AVs to serve intermodal rides as a priority and disincentivize serving non-intermodal rides.

To our knowledge, such a combination of taxes and subsidies has not been studied and optimized in large multimodal networks with multiple objectives. In most studies considering flat or link-based taxes, the price applies to travelers only (Gurumurthy *et al.*, 2020), or AVs only (Salazar *et al.*, 2018). Increasing AMoD fares typically represent a tax for travelers or a tax for AMoD operator, which passes it entirely on to its customers. When the regulator charges AVs only, the perfect market assumption usually justifies why the AMoD operator does not pass it on to its customers. This assumption is questionable. Using bi-level optimization allows accounting for the reaction of the AMoD operator to for-AV prices (Gao & Li, 2023, Dandl *et al.*, 2021). It deals with the trade-off between earning more money per ride on the one hand and losing customers with higher fares on the other hand. The operator may decide to split the tax and pass a part of it on to its customers if it improves its final profit. However, solving a bi-level optimization problem where the AMoD operator and the regulator respectively stand in the lower and upper levels introduces a hierarchy between these stakeholders. The AMoD operator is blind to the regulator's logic and only reacts to a certain regulation policy.

To overcome this limitation and explore all the possible splitting ratios, **we consider for-travelers and for-AVs prices as two separate decision variables for the regulator and optimize them regarding several objectives**. This way, the pricing scheme should grant and take money from the proper agents. Such a targeted strategy could spend the least regulating the system while deteriorating the least stakeholders indicators. As the AMoD fare scheme is assumed to be fixed and defined so that serving brings only a small profit, the total gain of the AMoD operator belongs to the objectives considered. Similarly, travelers' total travel cost and the equity in individual travel costs spatial distribution are considered goals for the regulator. More specifically, **we derive and analyze optimal policies regarding several sets of objective functions, including regulator's, AMoD operator's, and travelers' objectives**.

1.2.2 Refinement of some modeling bricks and their integration (C2 and C3)

1.2.2.1 State-of-the-art

An ideal modeling approach would account for a flexible intermodality, traffic dynamics, a detailed AMoD model, and coupling between supply and demand while remaining amenable to regulation measure optimization. Flexible intermodality is required because reducing the number of transportation alternatives introduces a bias in the analysis of the benefits of intermodal AMoD. Some mobility patterns emerging from individual usages of AVs may be missed by constraining travelers' choices to a limited set of modes and routes. One should consider all options to get all the cooperation/competition ranges between PT and AMoD. As large fleets of AVs may impact the traffic due to induced demand, empty relocations, and modal shifts, network loading dynamics should integrate the effect of AMoD endogenously. A detailed AMoD model accounts for the operational constraints inherent to ride-hailing systems, allows for a precise evaluation of the outputs for travelers, and better renders the sensitivity of the fleet to regulations. A more detailed evaluation of policies is possible when AMoD behavior and its interactions with customers are sharply modeled. Finally, since travelers continuously adapt their travel plans based on their experience and the detailed travel information available on the mobility system to make mode-route decisions, a realistic choice model and coupling between supply and demand are necessary.

As shown in Figure 1.5, several approaches have been proposed in the literature to model the components of a transportation system based on AMoD or intermodal AMoD. The two key components on the supply side are the AMoD and traffic flow models.

Traffic models The most straightforward approach ignores the impact of AMoD on traffic and considers exogenous congestion. Considering endogenous congestion requires a traffic model. Depending on the description scale, traffic flow models are usually classified into three categories (Figure 1.6).

- Microscopic models represent the traffic flow as a sequence of discrete vehicles interacting. Vehicle-to-vehicle interactions are described by car-following (Gipps, 1981) and lane-changing (Gipps, 1986) models. Huang *et al.* (2021), Zhu *et al.* (2018), Gueriau *et al.* (2020), Ban *et al.* (2021) use the microscopic traffic simulator SUMO (Lopez *et al.*, 2018) to simulate the deployment of AMoD. Microsimulation accounts for walk access details, transit facility geometries, and impacts on traffic of idling AVs during curbside pick-ups and drop-offs. However, their complexity and computational requirements make the approach incompatible with large-scale studies and regulation optimization.

- Mesoscopic models represent the traffic as continuous flows on the link level. Vehicle-to-vehicle interactions are implicit and described by phenomenological relationships such as in link-based queuing models (LWR - Lighthill & Whitham (1955), Richards (1956)) or point-queue models (Jin, 2015). Mesoscopic models miss the individual detailed behaviors of conventional cars and AVs but are less demanding in computing capacity. In the context of AMoD, analytical (Liu, 2018) and simulation (Hörl *et al.*, 2019, Militão & Tirachini, 2021) approaches have used mesoscopic models. Notably, the simulation platform MATSim, which is a state-of-the-art simulator for studying

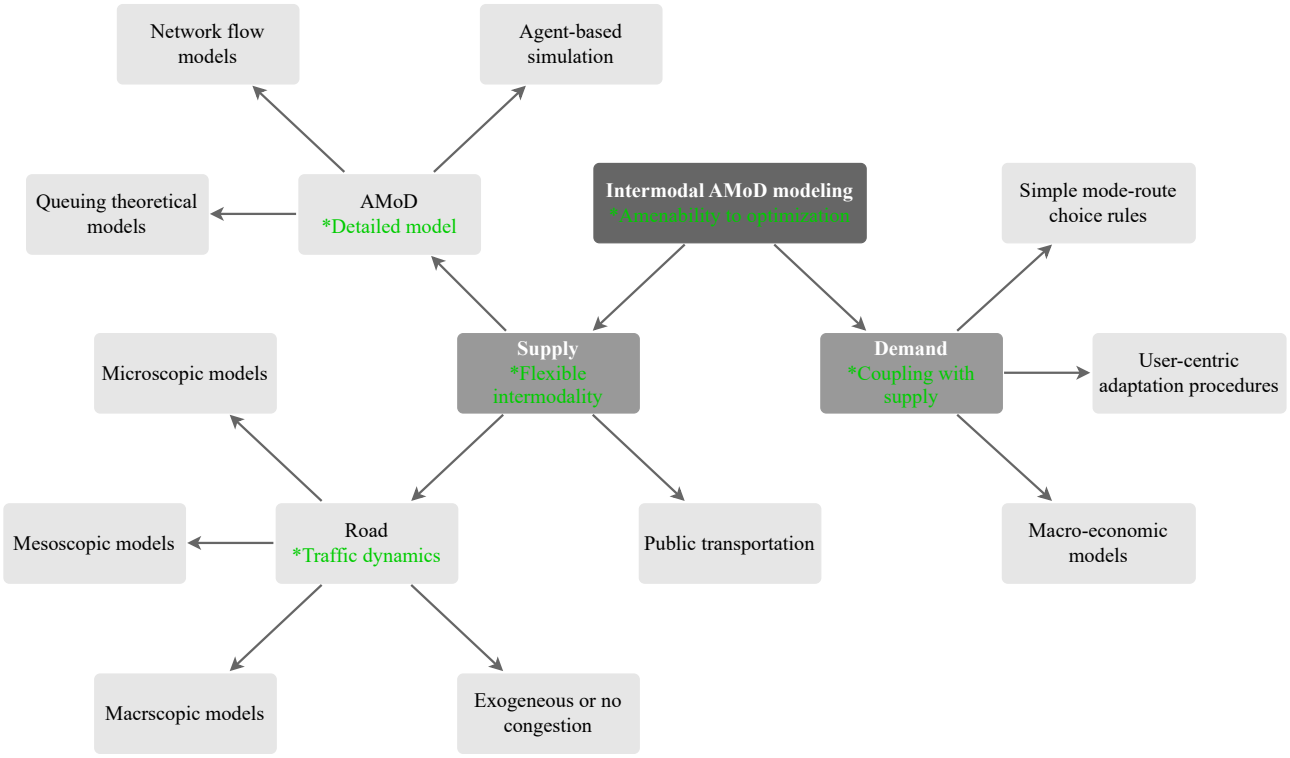


FIGURE 1.5 – *Approaches to model components of a transportation system based on intermodal AMoD. Green stars correspond to the desirable properties for tackling our research questions.*

AMoD and its integration in the urban transportation system, combines an incremental node model (Flötteröd & Rohde, 2011) and the Kinematic Wave Model (Newell, 1993) to represent traffic (Axhausen *et al.*, 2016, Agarwal & Lämmel, 2016). One can parameterize the Fundamental Diagram relationship between flow and density to account for the specific behavior of AVs, such as their reduced reaction times and shorter following distances compared to conventional cars (Simoni *et al.*, 2019).

- Macroscopic models represent the traffic at an aggregated level. The Macroscopic Fundamental Diagram (MFD) provides a relationship between network average density and flow in a given urban region (Mahmassani *et al.*, 1984, Geroliminis & Daganzo, 2007). MFD-based models describe traffic states per region and flow exchange between areas composing the city (Mariotte, 2018). They are computationally efficient, making the approach compatible with large-scale studies. However, the MFD form is challenging to estimate in the context of AMoD.

(A)MoD models Three main approaches have been used to model the operation of AMoD.

- Queuing theoretical models entail an exogenous dynamical stochastic process that generates AV ride requests at spatially localized queues (stations) (Pavone *et al.*, 2009, 2012, Pavone, 2010). AVs travel between locations to transport customers and relocate themselves. Pavone (2015) models the system as a closed Jackson network where stations are single-server queues and roads between them are infinite-server queues. Queuing theoretical models tackle the rebalancing problem in a car sharing system

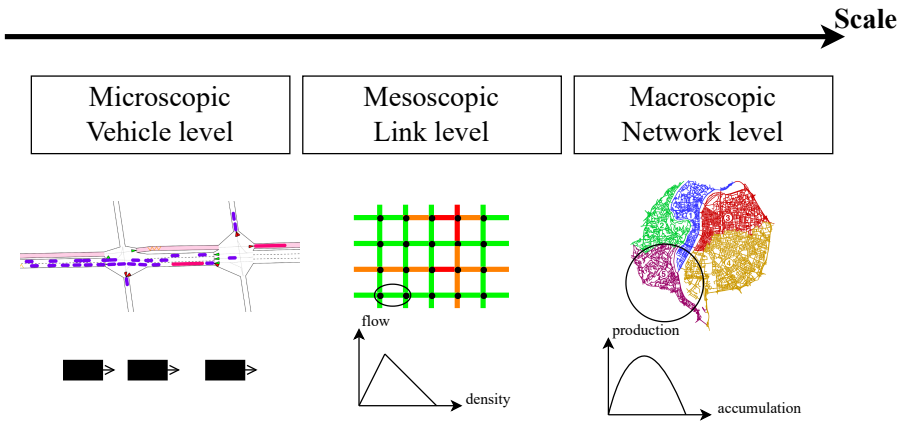


FIGURE 1.6 – *Scales of traffic description. Adapted from Mariotte (2018).*

with AVs efficiently (George, 2012, Zhang & Pavone, 2016, Zhang *et al.*, 2019). When stations spread over a continuous domain, the problem simplifies from controlling a finite number of localized queues into controlling a single spatially averaged queue. It leads to analytical expressions for some system metrics (e.g., availability of AVs or waiting time) and operational parameters (e.g., optimal fleet size). However, it removes the network's topology, rendering the coupling with other transportation modes and flexible intermodality inaccessible.

- Network flow models do not represent AVs and travelers individually but as flows on the transportation graph links. Time-invariant models reduce to static network flow problems (Salazar *et al.*, 2018, 2020) while time-variant models use time-expanded graphs (Zgraggen *et al.*, 2019). The approach is compatible with other modes of transportation and regulation measure optimization. It is consistent with any link-based traffic model in theory. However, in practice, simple functions, such as threshold, Bureau of Public Roads (BPR), or piecewise-affine approximation of BPR (Salazar *et al.*, 2019) functions provide interesting properties to the optimization problem tackled. Network flow models miss the individual behavior of AVs and travelers. Initially, they are not meant to design operational policies, but they enable planning studies. For example, Wollenstein-Betech *et al.* (2022) resort to a distributed algorithm to convert flows into recommended routes for AMoD users. The approach cannot render user-centric travel costs and metrics.

- Agent-based simulation disaggregates AVs and travelers. The approach reproduces interactions between travelers and the different modes with fidelity. Complex mode choice models (Rieser *et al.*, 2009, Axhausen *et al.*, 2010) and operational strategies Lin *et al.* (2018) can be incorporated to render user-centric and AMoD-centric objectives. Several agent-based simulation platforms exist (NetLogo, Swarm, MASON, AnyLogic). Some are dedicated to the transportation field and served in the context of AMoD or intermodal AMoD, such as POLARIS (Gurumurthy *et al.*, 2021), SimMobility (Basu *et al.*, 2018), and MATSim (Maciejewski *et al.*, 2016). Some authors have coded their custom platform (Wen *et al.*, 2018, Pöhler *et al.*, 2019, Yao *et al.*, 2020). Agent-based simulation can account for detailed AMoD modeling. However, it is poorly compatible with optimization algorithms in large-scale scenarios when the number of agents

is high, and the interaction rules are complex. The challenge for the agent-based simulation approach is to make it compatible with regulation optimization. Efficient and scalable algorithms for modeling AMoD with the proper level of detail and realism are necessary. Note that if the term "agent-based" is usually reserved for a specific range of simulation tools, it may be applied to a broader range of simulators as long as the environment and agents (entity sensing their environment and performing actions based on the information collected) are well-defined.

Demand models On the demand side, three approaches have been mainly adopted to study (intermodal) AMoD.

- Macro-economic models have been derived to be applied in the context of new mobility services deployment. Bourgeois (2017) has proposed a macro-economic model called OPTIREL to evaluate the relevancy of intermodal AMoD on specific territories. It adapts the traditional four-step model (Bonnel, 2004). The approach can provide an estimation of demand for a specific AMoD system. For example, Rifki *et al.* (2021) uses OPTIREL to evaluate the demand for an autonomous shuttle system providing the last mile service in an industrial area. The approach requires assuming AMoD's role and falls in constrained intermodality. It is only relevant to evaluate the performance and assets of an identified AMoD system in a bounded area compared to traditional mobility services.

- User-centric adaptation procedures rely on iterative assignment processes to mimic the adaptation of every traveler regarding her traveling experience. Noteworthy, this approach is compatible with agent-based simulation. For example, MATSim uses a co-evolutionary algorithm to reach the user equilibrium. Each traveler associates a score (or econometric utility) to an executed activity plan and can adapt this plan for the next day to maximize its utility. Replanning strategies aim to explore the research space correctly, while plan selection strategies aim to converge to equilibrium. This day-to-day equilibrium process is one of the most expensive components of simulators regarding computation time. This component is omitted in some studies to reduce the complexity of simulations. Instead, they consider simple mode and route choice rules.

- Defining mode and route choice rules do not necessarily render the relation between supply performances and associated demand. For example, some studies assume a constant demand for AMoD (supposing a given mode share or a certain proportion of personal car replacement), define precise rules (e.g., all travelers having their origin within a certain radius around a PT station choose to use PT), or use a mode choice model based on instantaneous travel costs.

The frontier between the approaches and models presented above regarding road traffic, AMoD operation, and demand-supply coupling is not clear and continuous. The choice for the proper mix of models depends on the research question and study case scale investigated.

The reader can find a more detailed state-of-the-art for shared autonomous vehicle services modeling in Narayanan *et al.* (2020) and for AMoD modeling (operation and control) in Zardini *et al.* (2022).

1.2.2.2 Contributions C2 and C3

Every approach proposed in the literature had to make some compromises between flexible intermodality, network loading dynamics, detailed AMoD model, coupling between supply and demand, and compatibility with optimization algorithms. In this thesis, we tackle two case studies in which the compromise and contributions differ.

In part II of the thesis, **we restrict the case study by considering an abstract corridor structured by two heavy linear transportation infrastructures**. In this type of configuration, substantial delays may emerge since the number of routes and transfer points between infrastructures is limited. Hence, the network loading dynamics and the coupling between supply and demand should be carefully addressed while maintaining amenability with optimization algorithms. **The contribution of this part holds in the corridor modeling and the efficient dynamic User Equilibrium (UE) resolution method**.

In part III of the thesis, **we treat a more generic network with numerous transportation alternatives**. In this configuration, congestion should spread across the network links. Delays should be smaller and less impact travelers' mode and route choices. However, the network geometry and the morning peak demand pattern special features are tough constraints for the operation of AMoD, including routing, matching, and relocating. Hence, **the first contribution of this part is the introduction and analysis of a detailed AMoD model representing the profit-oriented behavior of a fleet of AVs in an environment subject to for-AV taxes and subsidies**.

We represent the system with an agent-based framework. Since agent-based simulation requires substantial computational resources and produces outputs with no a priori mathematical properties, the second challenge addressed in this part is optimizing the regulation prices. To deal with it, we apply a recent methodology for efficient multi-objective optimization of time-expensive black-box functions. The **last contribution of this part lies in the integration of the AMoD operational strategy and the efficient multi-objective optimization method into an open-source, comprehensive agent-based simulation platform called Multimodal network Modeling and Simulation (MnMS)**.

1.3 Thesis outline

Figure 1.7 gives an overview of the chapters composing this thesis. The present chapter positions our work within the context and the state-of-the-art. It provides the keys for reading the manuscript. Parts II and III are independent and can be read in the order that fits the reader.

Part II tackles the corridor case study where the PT is a massive rapid transit (MRT) line. In chapter 2, we introduce a simple dynamic model subject to UE constraint to represent intermodal AMoD within this corridor while accounting for congestion. The model is described in three steps to properly understand UE principles, i.e., how UE settles and evolves. In chapter 3, we exploit this model to exhibit the link between some design parameters and the MRT-AMoD cooperation or competition. Sensitivity analysis allows us to understand better the circumstances of cooperation between the MRT and AMoD. We integrate our dy-

namic model into an optimization framework to optimize a new combination of push and pull design measures on a realistic scenario derived from the west of Lyon city (France).

Part III tackles the large urban area case study. In chapter 4, we propose an operational policy for an AV-centric large-scale AMoD model that renders the profit-oriented behavior of the AMoD operator and its sensitivity to monetary regulation measures. We analyze several policy variants in detail to specify their scope, i.e., the domain where the fleet is correctly managed and sensitive to a pricing scheme. In chapter 5, we formulate a multi-objective optimization problem for several monetary measures and objectives combinations. We present an efficient Multi-Objective Bayesian Optimization methodology to solve it. By integrating the AMoD operational policy and the solution method into a comprehensive agent-based simulation-multi-objective optimization framework, we assess the optimal regulation policies found and compare them to a baseline scenario where AMoD is absent and an unregulated scenario. We highlight the trade-offs between stakeholders' objectives by deriving the corresponding Pareto Fronts (PFs).

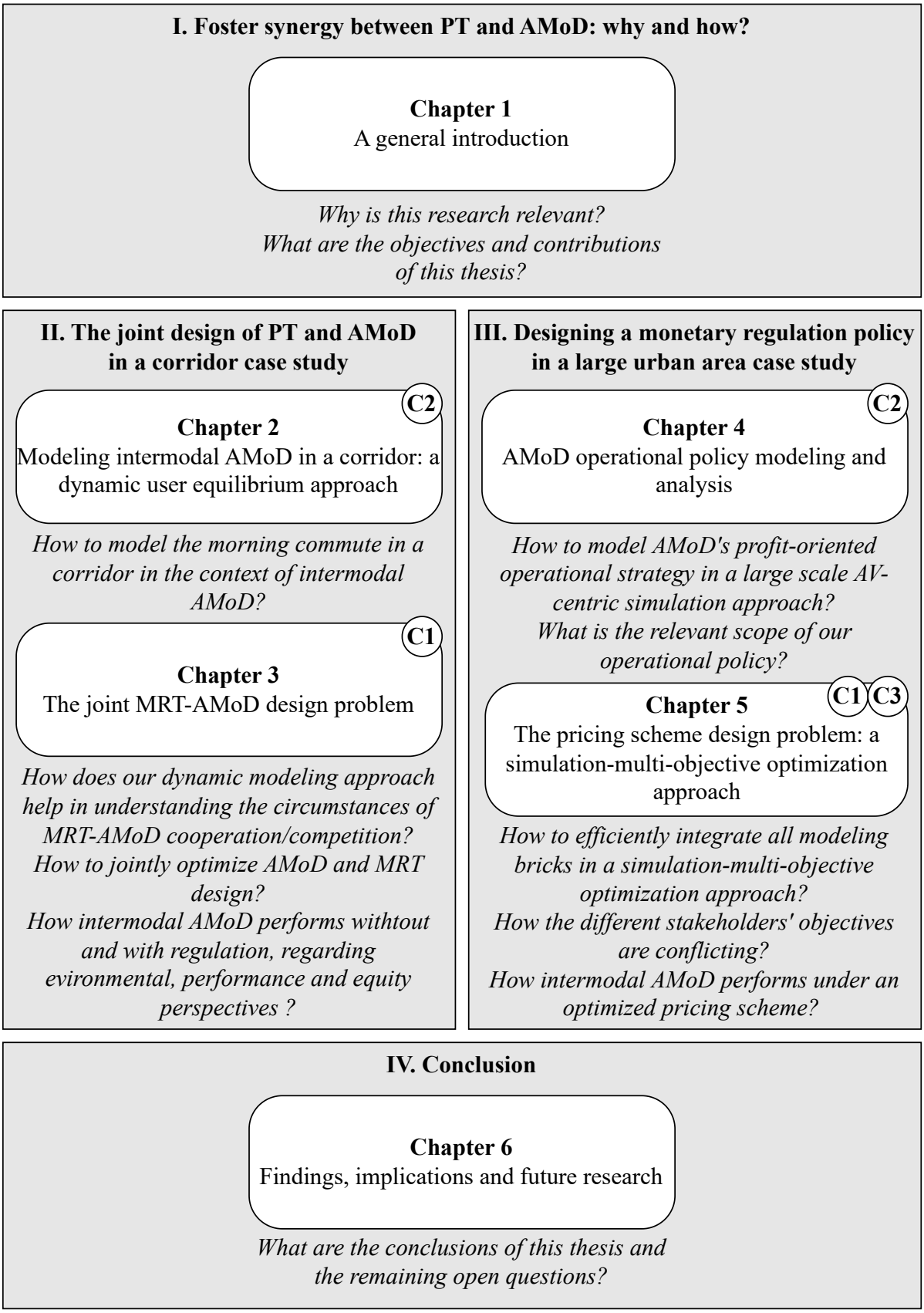


FIGURE 1.7 – *Thesis outline. Contribution C1, C2, C3, respectively relates to: AMoD-specific regulation policies building, the development or refinement of a modeling approach, brick, or methodology, the integration of modeling bricks and methodology together.*

1.4 Publication list

1.4.1 Peer-reviewed journal papers

- **Cortina, M.**, Chiabaut, N., Leclercq, L. (2023). Fostering synergy between transit and Autonomous Mobility-on-Demand systems: A dynamic modeling approach for the morning commute problem. *Transportation Research Part A: Policy and Practice*, 170:103638. https://doi.org/10.1016/j.tra.2023.103638. [Accepted]

1.4.2 Peer-reviewed conference proceedings

- **Cortina, M.**, Chiabaut, N., Leclercq, L. (2022). Dynamic modeling of morning commute in a monocentric corridor with autonomous vehicles, In *Transportation Research Board 101st Annual Meeting*, Washington DC, USA. [Accepted for poster]
- **Cortina, M.**, Chiabaut, N., Leclercq, L. (2023). An assignment-based dispatching heuristic for agglomeration-scale AMoD: sensitivity analysis of the demand prediction horizon on the benefits for supplier and customers, In *Transportation Research Board 102nd Annual Meeting*, Washington DC, USA. [Accepted for poster]
- **Cortina, M.**, Chiabaut, N., Leclercq, L. (2023). Equity of a batch-matching on horizon policy for Autonomous Mobility on Demand, In *24ème Congrès Annuel de la Société Française de Recherche Opérationnelle et d'aide à la décision*, Rennes, France. [Accepted for presentation]
- **Cortina, M.**, Khalesian, M., Leclercq, L. (2024). Multi-modal Traffic Management Optimization Using Gaussian Process and Pareto-based Multi-Objective Evolutionary Approach, In *Transportation Research Board 103rd Annual Meeting*, Washington DC, USA. [Accepted for poster]

Part II

The joint design of PT and AMoD in a corridor case study

2. Modeling intermodal AMoD in a corridor: a dynamic user equilibrium approach

Most elements of this chapter have been presented in:

- Cortina, M., Chiabaut, N., Leclercq, L. (2022). Dynamic modeling of morning commute in a monocentric corridor with autonomous vehicles, In *Transportation Research Board 101st Annual Meeting*, Washington DC, USA.

2.1 Motivation and objectives

The part II of this thesis focuses on the joint design of PT and AMoD as a regulation measure to foster their cooperation. As explained in chapter 1, an ideal modeling approach to evaluate the benefits of intermodal AMoD would account for: (i) the system dynamics, (ii) congestion, (iii) the presence of other transportation alternatives than intermodal AMoD, (iv) the mode and route choices of travelers, (v) operational capacities of the AV fleet.

Pure optimization and simulation approaches exist in the PT-(A)MoD design literature. Most optimization models are static in demand and travel times and do not account for congestion. Li & Quadrifoglio (2010) statically model a transportation corridor where a MoD system feeds an MRT line. They optimize the corridor configuration while ignoring potential congestion and other modes. They approximate the waiting time for an on-demand vehicle with a cycle time for each vehicle to serve its coverage zone following predefined rectilinear movements. Liu & Ouyang (2021) model an urban region as a union of square zones. Their model is static, does not account for congestion, and considers elementary rules to decide if a traveler journeys by PT+MoD (all inter-zonal trips) or MoD only (all intra-zonal trips). They use a queuing model to approximate the waiting time for MoD. Kumar & Khani (2022) also rely on a queuing model but tackle the case of real PT and road networks. They include assignment variables in a mixed integer non-linear program to capture travelers' behavior in the multimodal network. However, they do not consider other modes than PT+MoD: they formulate a pure route choice model with no mode choice. In contrast, Basciftci & Van Hentenryck (2023) consider the car alternative and induced demand. They formulate a bi-level optimization problem, including a simple mode choice model. However, they lack a MoD model by neglecting the waiting time for being picked up by an on-demand vehicle. Shan *et al.* (2021) make the same assumption on a corridor partitioned into several rectangle regions. They consider a time-variant demand, but travelers' choices between PT+AMoD and PT+Walk are predefined. MoD is used when no PT

station is available within the region, and walking is chosen otherwise. Wei *et al.* (2022) have proposed a dynamic model capturing MoD impact on road congestion and travelers' mode and route choices over several alternatives in a real network. They use a multinomial logit model for mode choice and a user equilibrium optimization model to assign MoD vehicles to routes across the roads. However, they address the PT design only, consider monomodal trips, and adopt exogenous MoD waiting times.

Simulation studies take into account the system's dynamics. As highlighted in chapter 1, simulation is an adequate tool to model the integration of PT and (A)MoD. However, complete simulation approaches, accounting for points (i) to (v), are barely compatible with design optimization. Point (iv) is particularly challenging. Indeed, computing the system's stable states in a dynamic environment, namely, the dynamic user equilibrium (DUE), is not trivial. In multimodal systems, user equilibrium refers to the situation when no user can reduce their own travel cost by unilaterally changing mode or route. Travelers are viewed as Nash agents competing on roads and modes capacities. Wardrop's first principle (Wardrop, 1952) defines user equilibrium in the traffic assignment problem by: "the journey times in all routes actually used are equal and less than those which would be experienced by a single vehicle on any unused route." Simulation approaches rely on computationally greedy convergence loops to compute the DUE (e.g., exact swapping or heuristics such as the method of successive average, evolutionary algorithms - Ameli (2019)). The PT-AMoD design optimization process, on top of dynamic traffic assignment, is costly, and its convergence is uncertain.

Few references tackle the PT-AMoD design problem with a simulation approach under the DUE. While Pinto *et al.* (2020) work in a cooperation context, Mo *et al.* (2021) question a competition scheme where AVs are unregulated and profit-oriented. Both studies consider real networks but ignore the competition with cars and the impact of AVs on traffic. Mo *et al.* (2021) only consider time-variant exogeneous congestion by deriving AVs and buses speeds from Google Map API. Moreover, simulation approaches suffer from the black box effect. They do not provide precise knowledge about how the PT and AMoD designs impact the way UE settles and evolves.

Table 2.1 summarizes the modeling choices for the works cited above. The state-of-the-art lacks methods compatible with PT-AMoD design optimization to solve and fully understand the DUE. Our work partially fills this gap by proposing a simple but dynamic model for computing the DUE in the context of the morning commute in a corridor. We focus on a restrictive study case regarding network and intermodality to fully understand how UE settles and evolves. Our model exhibits the links between PT-AMoD design parameters, congestion dynamics, and cooperation-competition between PT and AMoD. Its computational efficiency makes it easy to couple with a design optimization heuristic.

The main objective of this chapter is to describe our dynamic model. To clarify the DUE principles, we present the model in three steps. The first step presents the simplest version of the model, which holds within a linear corridor under the toughest assumptions in terms of travelers' origins distributions and access times. The second step presents a pseudo-dynamic version of the model, within a square corridor, under a strong assumption in terms of waiting time for AMoD. The last step presents the final dynamic model exploited in chapter 3 for the PT-AMoD design problem.

Another objective of this chapter is to present an equivalent static version of the model, used for benchmarking in chapter 3.

Reference	Approach	Area	Dynamics	Congestion	Alternative	Travelers choices	(A)MoD model
Li & Quadrifoglio (2010)	Optim.	Corridor	✗	✗	✗	✗	Cycle time approxim.
Liu & Ouyang (2021)	Optim.	Area partitioned in square regions	✗	✗	MoD only	Predefined	Queuing model
Kumar & Khani (2022)	Optim.	Real network	✗	✗	✗	Route only	Queuing model
Basciftci & Van Hentenryck (2023)	Optim.	Real network	✗	✗	Car	Mode only	✗
Shan *et al.* (2021)	Optim.	Corridor partitioned in rectangle regions	Time-variant demand	✗	Walk+PT	Predefined	✗
Wei *et al.* (2022)	Optim.	Real network	✓	✓	Monomodal trips only (PT, MoD, car, other)	Multinomial logit for mode, UE for routes	Exogen. MoD waiting time, idle moves reconstruction
Pinto *et al.* (2020)	Simul.	Real network	✓	✗	Walk+PT, AMoD only	Multinomial logit for mode, UE for routes	Model-based approxim.
Mo *et al.* (2021)	Simul.	Real network	✓	Exogen.	Walk+PT, AMoD only	Mixed logit for mode, UE for routes	Event-based operational policy simulated
Ours	Simple simul.	Corridor	✓	Endogen. no AVs-cars interactions	Walk+PT, Car	Deterministic UE for mode and route	Model-based approxim.

TABLE 2.1 – *Comparison of the modeling approaches for the PT-AMoD design problem (✓= taken into account, ✗= not accounted).*

2.2 Notations for this chapter

TABLE 2.2 – *Notations for chapter 2.*

Notation	Definition
	Linear corridor model
B	Corridor length
M	Number of collectors along the corridor (except destination collector c_0)
c_k	k^{th} collector of the corridor ($k \in [\![0, M]\!]$)
x_k	Location of collector c_k ($k \in [\![0, M]\!]$)
$D_{c_k}(t)$	Curve of departures from c_k ($k \in [\![1, M]\!]$)
v_{r}	Cruising speed of the MRT
T_{d}	Dwell time at a MRT station
h	Headway of MRT service during peak hours in seconds
T_{s}	Fixed AMoD service time
u	Free-flow speed on freeway
μ_k	Capacity of bottleneck associated with c_k, it also designates this bottleneck in the text ($k \in [\![0, M]\!]$)
T_{f}	Time for the drop-off maneuver
$w_k(t)$	Waiting time at bottleneck μ_k at t ($k \in [\![0, M]\!]$)
$A_k(t)$	Cumulative number of travelers arrived at bottleneck μ_k by t ($k \in [\![0, M]\!]$)
t_k	Time at which congestion starts at bottleneck μ_k ($k \in [\![0, M]\!]$)
T_k^{c}	Free-flow travel time of travelers departing from c_k by mode **c** ($k \in [\![1, M]\!]$)
$T_{k,p}^{\text{a}}$	Free-flow travel time of travelers departing from c_k by mode **a** with a transfer at c_p ($k \in [\![1, M]\!]$, $p \in [\![1, k]\!]$)
T_k^{r}	Travel time of travelers departing from c_k by mode **r** ($k \in [\![1, M]\!]$)
$\tau_k^{\text{c}}(t)$	Travel time of travelers departing from c_k using mode **c** by t ($k \in [\![1, M]\!]$)
$\tau_{k,p}^{\text{c}}(t)$	Travel time of travelers departing from c_k using mode **a** with a transfer at c_p by t ($k \in [\![1, M]\!]$, $p \in [\![1, k]\!]$)
Δ_p^{a}	Additional travel time experienced in free-flow conditions by choosing mode **a** with a transfer at c_p instead of **c** ($p \in [\![1, M]\!]$)
Δ_k^{r}	Additional travel time experienced in free-flow conditions by choosing mode **r** instead of **c** ($k \in [\![1, M]\!]$)
$A_k^{\text{th}}(t)$	Theoretical arrival curve at bottleneck μ_0 ($k \in [\![1, M]\!]$)
$A^{\text{th}}(t)$	Global theoretical arrival curve at bottleneck μ_0
$A_k^{\text{r}}(t)$	Curve of the cumulative count of travelers arriving/departing at/from c_k by mode **r** ($k \in [\![1, M]\!]$)
	Pseudo-dynamic model
W	Corridor width
(x_i, y_i)	Traveler i's origin coordinates
$v_{\text{st}}, v_{\text{w}}$	Speed of cars in streets, speed of pedestrians
t_i^{req}	Time at which traveler i sends a request for AV

Continue on the next page

TABLE 2.2 – *Notations for chapter 2 (continued).*

Notation	Definition
T_{s}, $T_{\mathrm{s}}(t_i^{\mathrm{req}})$	AMoD service time function, service time experienced by traveler i
Ω_k^{w}	Walking attraction area of c_k ($k \in [\![1, M]\!]$)
$X_{k,k+1}^{\mathrm{w}}$	Frontier between Ω_k^{w} and Ω_{k+1}^{w} ($k \in [\![1, M-1]\!]$)
Ω_k^{c}	Vehicle attraction area of c_k ($k \in [\![1, M]\!]$)
$X_{k,k+1}^{\mathrm{c}}$	Frontier between Ω_k^{c} and $\Omega_{k+1}^{\mathrm{c}}$ ($k \in [\![1, M-1]\!]$)
T_i^{c}	Free-flow travel time of traveler i by mode **c**
$T_{i,p}^{\mathrm{a}}$	Free-flow (neither waiting for AMoD nor waiting at bottleneck μ_p are included) travel time of traveler i by mode **a** with a transfer at c_p (c_k being i's vehicle access collector, $p \in [\![1, k]\!]$)
T_i^{r}	Travel time of traveler i by mode **r**
s_k	Spacing between c_k and c_{k+1}
t_i^0	Theoretical arrival time of traveler i at c_0
t_i^{b}	Time at which traveler i arrives at bottleneck μ_p when they choose mode **a** (c_k being i's vehicle access collector, $p \in [\![1, k]\!]$)
τ_i^{c}	Travel time of traveler i by mode **c**
$\tau_{i,p}^{\mathrm{a}}$	(Travel time of traveler i by mode **a** with at transfer at c_p (c_k being i's vehicle access collector, $p \in [\![1, k]\!]$)
Δ_i^{r}	Additional travel time experienced by traveler i when they take mode **r** instead of mode **c** in free-flow conditions
$\Delta_{i,p}^{\mathrm{a}}$	Free-flow additional travel time experienced by traveler i when they take mode **a** with a transfer at c_p instead of free-flow mode **c** (c_k being i's vehicle access collector, $p \in [\![1, k]\!]$)
$D_{c_k}^{\mathrm{th}}(t)$	Theoretical curve of departures from c_k ($k \in [\![1, M]\!]$)
	Dynamic model
m	Number of AVs in the fleet
t_i^{dep}	Time at which traveler i departs from home ($t_i^{\mathrm{dep}} = t_i^{\mathrm{req}}$)
$\widehat{T}_{\mathrm{s}}$	Effective service time profile
Δt	Time for the fleet to receive the m last request
E_1, E_2, E_3	Expected pick up time for the mean request, expected running time for the mean request, expected waiting time at bottleneck for the mean request
	Static model
$G = (V, E)$	Graph representing the corridor
N	Number of traveler pools
O_i	Abstract origin aggregating several commuters ($i \in [\![1, N]\!]$)
D, C_k^{c}, C_k^{a}, C_k^{r}	Destination, on-ramp, station and transfer nodes in G ($k \in [\![1, M]\!]$)
α_{c}, β_{c}, α_{a}, β_{a}	BPR parameters respectively associated with freeway final off-ramp and drop-off delay generator points
$\nu, \tau(\mathrm{edge})$	Traffic volume, travel time on edge of G

2.3 A linear corridor model

The linear corridor model is the simplest version. It describes how the flows of travelers distribute over the itineraries, focusing on the two major transportation infrastructures composing the corridor while ignoring accessibility issues. It allows defining and illustrating the DUE principles that hold in the two other versions of the model.

2.3.1 Problem setting and assumptions

Figure 2.1 presents the assumptions made to tackle this problem and the associated parameters. The following sub-sections describe each of them.

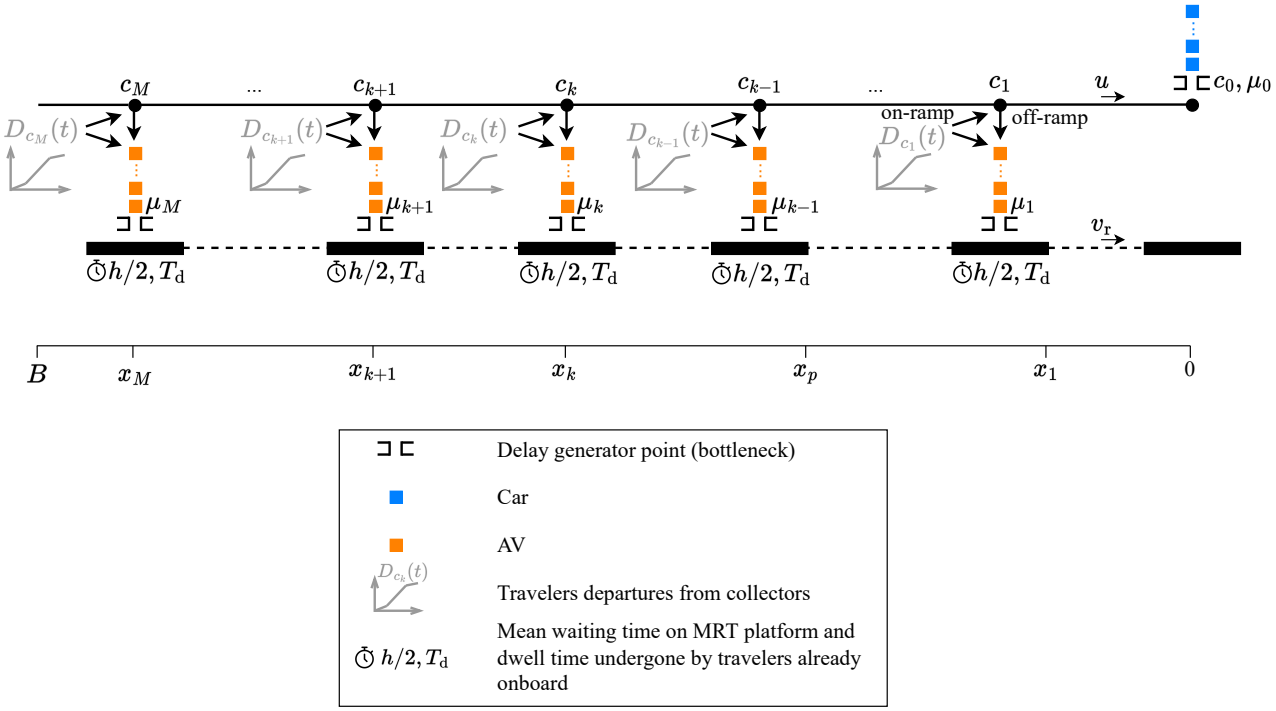


FIGURE 2.1 – *The linear corridor model.*

2.3.1.1 Multimodal corridor infrastructure

Let us consider the part of a monocentric city surrounding a linear transportation axis composed of a freeway and an MRT system (e.g., suburban rail, subway, or tram). This corridor stretches from the city boundary B to the Center of the Business District (CBD) located at $x = 0$. Both the freeway and MRT line are assumed to overlap at $y = 0$ and to be connected through $M + 1$ collectors c_k located along the corridor at x_k, $k \in [\![0, M]\!]$. A collector is composed of a freeway on-ramp, an off-ramp, and an MRT station with its drop-off spots. Collector c_0 corresponds to the terminus station and the off-ramp leading to CBD. Figure 2.2 represents the corridor.

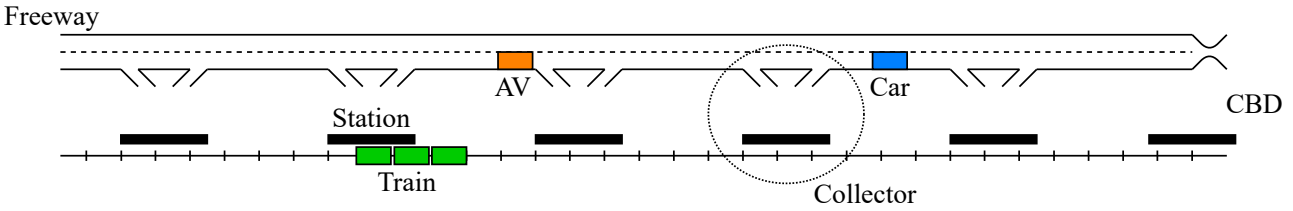


FIGURE 2.2 – *Corridor representation. The freeway and MRT line are the two infrastructures in the corridor and are connected through collectors.*

2.3.1.2 Demand spatial and temporal profiles

Commuters depart from home and want to join the CBD. There are many origins for one destination. A Dirac-like distribution of travelers' origins along x axis is considered: travelers' origins are located at collectors' abscissa x_k, $k > 0$. Consequently, the travel time to access the freeway or the MRT line is null. Section 2.4 relaxes this assumption.

Three trip generation rates are defined at each collector. They correspond to three distinct phases of the morning commute, as shown in Figure 2.3: low loading, high loading, and unloading phases. We note the curve of departures from c_k $D_{c_k}(t)$ for $k \in [\![1, M]\!]$. Indeed, for the DUE resolution in the linear corridor model, we do not consider the travelers departing directly from c_0.

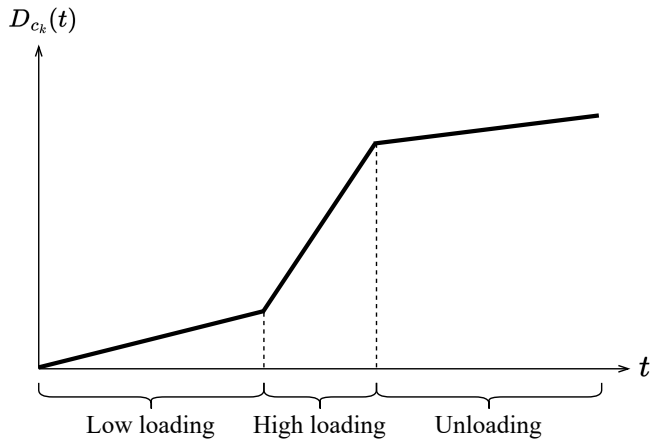


FIGURE 2.3 – *Departures of travelers from c_k. Three trip generation rates are considered over time at each collector.*

2.3.1.3 Deterministic mode and route choice based on travel time

Three transportation alternatives exist for commuters: car only (**c**), MRT only (**r**), and AV+MRT (**a**). Each traveler chooses the mode and route that provides a minimal travel time. Backward movements on the freeway are prohibited. A traveler departing from c_k has to choose between $k + 2$ itineraries:

- one corresponding to **c**: traveler enters the freeway driving their own car and stays on it until c_0,

- one corresponding to **r**: traveler boards the MRT at c_k's station and rides until the terminus,
- k corresponding to **a**: traveler boards an AV and chooses to transfer to the MRT at c_p, where $p \in [\![1, k]\!]$. Indeed, AMoD operates within the corridor only, and AVs are not allowed to drop commuters off directly in the CBD.

2.3.1.4 MRT operation

All stations belong to the same MRT line: a train stops at each one. This line is assumed to have an unlimited passenger capacity, i.e., there is no left-behind on platforms because of saturated trains. Let us neglect the MRT rolling stock kinematics. Since the wasted time due to limited acceleration and braking capacities is ignored, no minimal spacing between two stations is required. The cruising speed of trains is v_r. We also neglect the relation between the number of travelers boarding at a station and the dwell time. This assumption is consistent with the unsaturated line hypothesis. Dwell time is therefore constant, equal for all stations, and noted T_d. The more collectors there are, the longer the total dwell time, and the lower the commercial speed of the MRT line. The design parameter M impacts the MRT line travel time to the CBD. Line headway h (in seconds) is considered stable during the morning commute. The time a traveler needs to wait at a station is approximated by $\frac{h}{2}$ (Fu *et al.*, 2012).

2.3.1.5 AMoD operation

AMoD comprises one fleet of single-seat AVs operating on the whole corridor with the exception of the CBD. The waiting time undergone by a traveler before being picked up by an AV is called the service time and assumed fixed here. Note that booking is disabled for AMoD: a commuter sends a request at the moment when they depart from home, and undergoes from then a fixed service time T_s. This assumption is valid as long as the number of vehicles operating in the fleet is sufficient, i.e. when the service has been correctly dimensioned in advance. This assumption is removed in the dynamic model.

2.3.1.6 Delay generator points

Localized delay generator points exist in this network. A vehicle runs on the freeway at speed u in free-flow conditions. It may be caught in congestion due to a single capacity reduction point. The queuing delay experienced by a car exiting the freeway at c_0 is due to a demand rate higher than the destination off-ramp fixed capacity μ_0.

Moreover, queues form at AV-to-MRT transfer points due to the fixed number of stopping spots and the fixed time T_f needed for the drop-off maneuver (for an AV to park at a drop-off spot and for the passenger to leave the vehicle in security). These delays are the only ones AVs face. The continuum approximation of the drop-off times comes down to modeling AV-to-MRT transfer capacity at a station by a fixed value μ_k. In practice, adding drop-off spots will increase this capacity.

A point-queue model provides traffic delays. Introduced by Vickrey (1969), this model first dealt with the departure time choice of commuters on a single bottleneck-constrained

one-to-one corridor (Arnott *et al.*, 1990). It has been extended to many-to-one (Fosgerau & de Palma, 2012), multiple bottlenecks (Akamatsu *et al.*, 2015), and multimodal corridor (Sean Qian & Michael Zhang, 2011, Wu & Huang, 2014, Chiabaut *et al.*, 2018). It is one of the simplest models accounting for congestion dynamics. It can deal with aggregated (flow) and disaggregated (individual travelers) points of view, allowing the analytical and numerical resolution of the DUE. Delays given by the point-queue model are equivalent to the LWR model with spreading congestion. Vertical queues allow modifying the corridor configuration while preventing disturbance of upstream off-ramps by spillback congestion. Cars and AVs flows do not interact in our model. This assumption is reasonable when c_1 is sufficiently far from c_0 compared to queue length, or AVs can run on a dedicated lane on the freeway.

Considering that traffic obeys a first-in-first-out (FIFO) rule at each bottleneck μ_k, $k \in [\![0, M]\!]$, the waiting delay can be formulated by:

$$w_k(t) = \max\left(0, \frac{A_k(t) - A_k(t_k)}{\mu_k} - (t - t_k)\right) \tag{2.1}$$

where $A_k(t)$ denotes the cumulative number of travelers arrived at bottleneck μ_k by t, t_k represents the time at which congestion starts at bottleneck μ_k. For $k \in [\![1, M]\!]$, $w_k(t)$ includes T_f, the time for an AV to park at a drop-off spot and for the passenger to leave the vehicle in security. The application of the point-queue model allows writing delays as simple functions of time.

2.3.1.7 Travel times

One can express the travel times on each possible itinerary for the travelers departing from c_k. The free-flow travel time by mode **c** is:

$$T_k^\mathrm{c} = \frac{x_k}{u} \tag{2.2}$$

The free-flow travel time by mode **a** with an AV-to-MRT transfer at c_p ($p \in [\![1, k]\!]$) is:

$$T_{k,p}^\mathrm{a} = T_\mathrm{s} + \frac{x_k - x_p}{u} + \frac{h}{2} + \frac{x_p}{v_\mathrm{r}} + (p-1)T_\mathrm{d} \tag{2.3}$$

The travel time by mode **r** is time-invariant:

$$T_k^\mathrm{r} = \frac{h}{2} + \frac{x_k}{v_\mathrm{r}} + (k-1)T_\mathrm{d} \tag{2.4}$$

Outside of free-flow conditions, the travel time by mode **c** is time-variant:

$$\tau_k^\mathrm{c}(t) = T_k^\mathrm{c} + w_0(t) \tag{2.5}$$

where t is the arrival time of traveler at bottleneck μ_0.
Similarly, outside of free-flow conditions, the travel time by mode **a** with a transfer at c_p ($0 < p \leq k$) is time-variant:

$$\tau_{k,p}^\mathrm{c}(t) = T_{k,p}^\mathrm{a} + w_p(t) \tag{2.6}$$

where t is the arrival time of traveler at bottleneck μ_p.

The difference $T^{\mathrm{a}}_{k,p} - T^{\mathrm{c}}_k$ only depends on p, not on k since T_{s} is fixed for all travelers of the corridor. Then, we note $\Delta^{\mathrm{a}}_p = T^{\mathrm{a}}_{k,p} - T^{\mathrm{c}}_k$ the additional travel time experienced in free-flow conditions by choosing mode **a** with a transfer at c_p instead of mode **c**. We note $\Delta^{\mathrm{r}}_k = T^{\mathrm{r}}_k - T^{\mathrm{c}}_k$ the additional travel time experienced in free-flow conditions by choosing mode **r** instead of mode **c**.

2.3.1.8 Equilibrium

We study the system at equilibrium. We assume the system conforms to the UE defined by Wardrop's first principle. If this principle has been introduced in the context of continuous flows of vehicles for the route choice problem, it is applied here on continuous flows of travelers for the mode-route choice problem. No one can reduce their travel times by unilaterally choosing another mode-route of the same origin-destination pair. Note that departure times are parameters of the problem, not decision variables.

2.3.2 Dynamic user equilibrium solution

In this section, we present the DUE resolution process. It is inspired by Laval (2009). The original method performs in a one-to-one corridor with cars only. We developed an extended method to deal with a many-to-one corridor with three modes. The DUE principles are defined and graphically illustrated based on two examples.

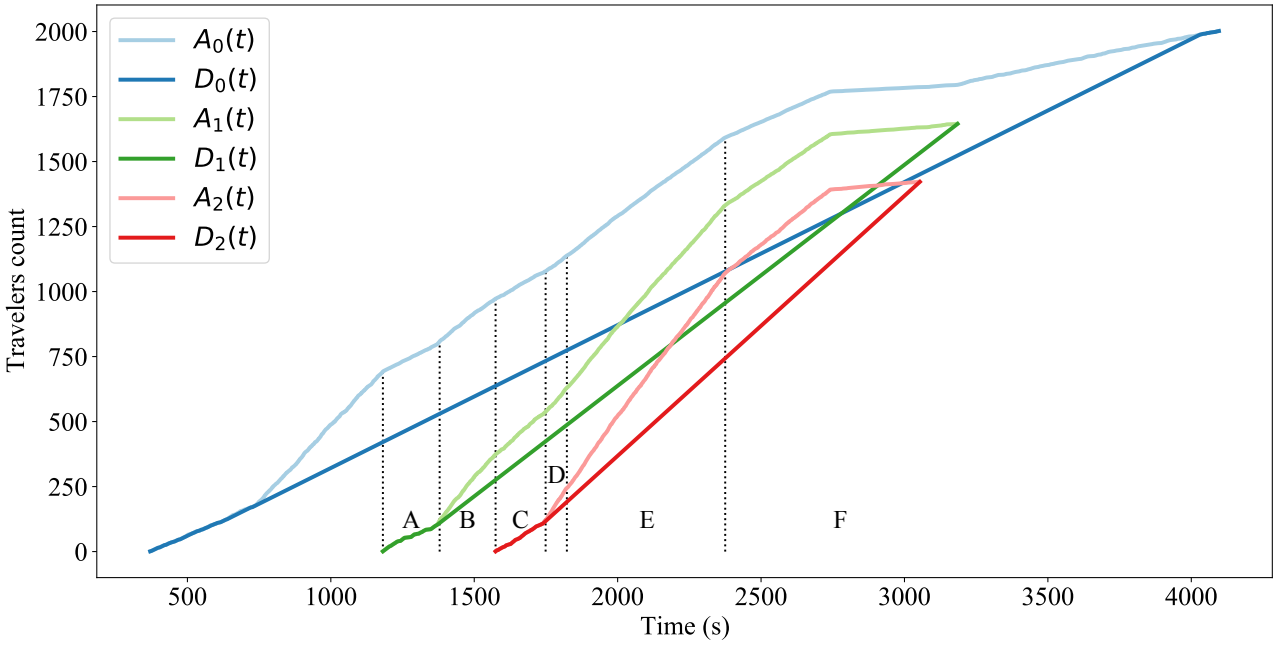
The free-flow speed on the freeway is assumed to be greater than the MRT cruising speed ($v_{\mathrm{r}} < u$). It means that all travelers initially prefer mode **c**: $\Delta^{\mathrm{a}}_p \geq 0$ for $p \in [\![1, M]\!]$, and $\Delta^{\mathrm{r}}_k \geq 0$ for $k \in [\![1, M]\!]$. Moreover, collectors abscissa are distinct, so that the free-flow additional travel times on **a** diversion routes are monotonically increasing from downstream to upstream collectors ($\Delta^{\mathrm{a}}_1 < ... < \Delta^{\mathrm{a}}_M$). We note $A^{\mathrm{th}}_k(t)$ c_k's theoretical arrival curve at bottleneck μ_0, which is computed by assuming that all travelers departing from c_k choose mode **c**. It is obtained by translating $D_{c_k}(t)$ by x_k/u. A^{th}(t) is the global theoretical arrival curve at bottleneck μ_0. It is the sum of all the $A^{\mathrm{th}}_k(t)$. The same translation is applied to $A_k(t)$, $w_k(t)$, and $\tau^{\mathrm{a}}_{k,p}(t)$ ($k \in [\![1, M]\!]$) to align all curves in time with bottleneck μ_0. This translation does not impact the derivation of flows dynamics.

2.3.2.1 DUE with modes c and a available

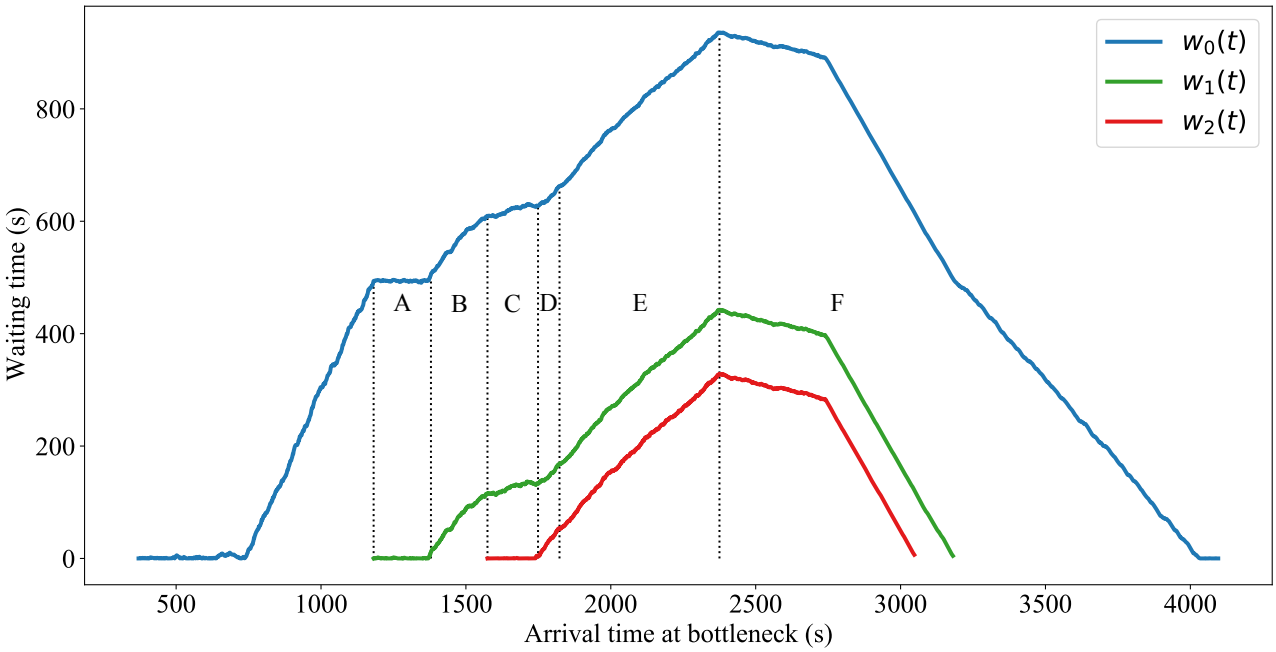
In this first example, only modes **c** and **a** are available. We describe the DUE on an example containing only two collectors and then generalize to any M value.

Example with $M = 2$ Figure 2.4 presents the DUE chronologically described below.

1. Initially, every commuter chooses **c**. Congestion on the freeway starts as soon as $\dot{A}^{\mathrm{th}}(t) > \mu_0$ where $A^{\mathrm{th}}(t) = A^{\mathrm{th}}_1(t) + A^{\mathrm{th}}_2(t)$ and dot on top of a variable denotes time differentiation. A queue forms and grows until $w_0(t) = \Delta^{\mathrm{a}}_1$. As c_1 is the most downstream collector before the destination, it is accessible to everyone. All travelers have a new alternative to c_0 itinerary.



(a) *Arrival and departure curves at μ_0, μ_1 and μ_2 bottlenecks.*



(b) *Waiting times at μ_0, μ_1 and μ_2 bottlenecks.*

FIGURE 2.4 – *DUE solution for the linear corridor model with M=2, modes **c** and **a** available. Diversion pattern **a**-div1 takes place during period A, and **a**-div2 during period B. Bottleneck μ_2 falls out of sync with downstream bottlenecks for period C and re-synchronizes for period D. Pattern **a**-div2 continues on period E and F until all queues vanish. Note that all the curves are translated in time to be aligned with μ_0 bottleneck and allow a better highlight of the diversion patterns.*

2. According to Wardrop's first principle, travel times of all routes used should be equal: $\tau_k^{\mathrm{c}}(t) = \tau_{k,1}^{\mathrm{a}}(t)$, $\forall k \in \{1,2\}$. If $\dot{A}^{\mathrm{th}}(t) \leq \mu_0 + \mu_1$, mode **a** starts to be used by both flows of travelers with a transfer at c_1, but μ_1 remains uncongested. So travel times

equality can be derived into $\dot{w}_0(t) = 0$. Replacing w_0 by its expression given by 2.1 leads to: $\dot{A}_0(t) = \mu_0$ which characterizes **a** diversion pattern 1 (**a**-div1). See period A in Figure 2.4.

3. As soon as $\dot{A}^{\text{th}}(t) > \mu_0 + \mu_1$, bottleneck μ_1 starts to be congested, and **a** diversion pattern 2 (**a**-div2) is observed (period B). Applying Wardrop's first principle again leads to $\dot{w}_0(t) = \dot{w}_1(t)$: waiting times at bottlenecks μ_0 and μ_1 increase at the same pace. As all travelers pass through c_0 and c_1, we have $\dot{A}^{\text{th}}(t) = \dot{A}_0(t) + \dot{A}_1(t)$. It follows:

$$\begin{cases} \dot{A}_0(t) = \dfrac{\mu_0}{\mu_0 + \mu_1} \dot{A}^{\text{th}}(t) \\ \dot{A}_1(t) = \dfrac{\mu_1}{\mu_0 + \mu_1} \dot{A}^{\text{th}}(t) \end{cases} \tag{2.7}$$

until $w_0(t) = w_1(t) + \Delta_1^{\text{a}} = \Delta_2^{\text{a}}$.

4. At this time, we have $\mu_0 + \mu_1 < \dot{A}^{\text{th}}(t) \leq \mu_0 + \mu_1 + \mu_2$. Mode **a** with a transfer at c_2 starts to be a competitive option. This new diversion alternative is only accessible to travelers departing from c_2. Two cases are possible:

   - If $\dot{A}_1^{\text{th}}(t) \leq \mu_0 + \mu_1$, the flow of travelers from c_2 still use bottlenecks μ_0 and μ_1. Waiting times $w_0(t)$ and $w_1(t) + \Delta_1^{\text{a}}$ are maintained equal to Δ_2^{a}. It is a **a**-div1 pattern where μ_0 and μ_1 work at capacity while μ_2 absorbs the surplus of demand coming from c_2.

$$\begin{cases} \dot{A}_0(t) = \mu_0 \\ \dot{A}_1(t) = \mu_1 \\ \dot{A}_2(t) = \dot{A}^{\text{th}}(t) - \mu_0 - \mu_1 \quad \leq \mu_2 \end{cases} \tag{2.8}$$

   - If $\dot{A}_1^{\text{th}}(t) > \mu_0 + \mu_1$, **a**-div2 is maintained for the flow of travelers departing from c_1 only, while mode **a** with a transfer at c_2 is chosen by all travelers departing from c_2. Waiting time at bottleneck μ_2 falls out of sync with other bottlenecks: $w_0(t) = w_1(t) + \Delta_1^{\text{a}} > \Delta_2^{\text{a}}$ and $w_2(t) = 0$ (period C).

$$\begin{cases} \dot{A}_0(t) = \dfrac{\mu_0}{\mu_0 + \mu_1} \dot{A}_1^{\text{th}}(t) \\ \dot{A}_1(t) = \dfrac{\mu_1}{\mu_0 + \mu_1} \dot{A}_1^{\text{th}}(t) \\ \dot{A}_2(t) = \dot{A}_2^{\text{th}}(t) \end{cases} \tag{2.9}$$

5. If global demand is still increasing, we get to a time when $\dot{A}^{\text{th}}(t) > \mu_0 + \mu_1 + \mu_2$. Again, several cases arise depending on the synchronization status between bottlenecks and demand per origin collector.

   - If bottlenecks are all synchronized ($w_0(t) = w_1(t) + \Delta_1^{\text{a}} = \Delta_2^{\text{a}}$), and $\dot{A}_1^{\text{th}}(t) \leq$

$\mu_0 + \mu_1$ or $\frac{\dot{A}_1^{\text{th}}(t)}{\mu_0+\mu_1} < \frac{\dot{A}_2^{\text{th}}(t)}{\mu_2}$, a global **a**-div2 is observed:

$$\begin{cases} \dot{A}_0(t) = \dfrac{\mu_0}{\mu_0+\mu_1+\mu_2}\dot{A}^{\text{th}}(t) \\ \dot{A}_1(t) = \dfrac{\mu_1}{\mu_0+\mu_1+\mu_2}\dot{A}^{\text{th}}(t) \\ \dot{A}_2(t) = \dfrac{\mu_2}{\mu_0+\mu_1+\mu_2}\dot{A}^{\text{th}}(t) \end{cases} \tag{2.10}$$

- If bottleneck μ_2 is not synchronized with others ($w_0(t) = w_1(t) + \Delta_1^{\text{a}} > \Delta_2^{\text{a}}$), and $\dot{A}_1^{\text{th}}(t) \leq \mu_0 + \mu_1$ or $\frac{\dot{A}_1^{\text{th}}(t)}{\mu_0+\mu_1} < \frac{\dot{A}_2^{\text{th}}(t)}{\mu_2}$, **a**-div2 continues for the flow of travelers departing from c_1 while mode **a** with a transfer at c_2 is chosen by the other flow until re-synchronization. Re-synchronization of bottlenecks group $\{\mu_0,\mu_1\}$ with bottleneck μ_2 takes place when $w_0(t) = w_1(t) + \Delta_1^{\text{a}} = w_2(t) + \Delta_2^{\text{a}}$, at the end of period D, from when a global **a**-div2 pattern starts (period E).
- If $\frac{\dot{A}_1^{\text{th}}(t)}{\mu_0+\mu_1} > \frac{\dot{A}_2^{\text{th}}(t)}{\mu_2}$, whatever the synchronization state of bottlenecks, **a**-div2 continues for the flow of travelers departing from c_1 while mode **a** with a transfer at c_2 is chosen by the other flow. Re-synchronization is not going to occur for now.

6. When global demand decreases, a bottleneck ceased to be used as soon as its queue vanishes. If bottlenecks are synchronized, **a**-div2 continues (period F). If they are not, re-synchronization takes place during the decrease in demand.

Any M value During the loading phase, **a**-div1 and **a**-div2 emerge from groups of synchronized bottlenecks. During the unloading phase, **a**-div2 maintains until the end of the morning commute. The DUE can be computed following a set of simple rules.

- Bottleneck μ_0 is used from the beginning of the morning commute.
- Bottleneck μ_{p+1} starts to be used as soon as $w_p(t) + \Delta_p^{\text{a}} = \Delta_{p+1}^{\text{a}}$.
- Two bottlenecks μ_p and μ_q ($p > 0$ and $q > 0$) are said synchronized when $w_p(t) + \Delta_p^{\text{a}} = w_q(t) + \Delta_q^{\text{a}}$. Bottleneck μ_p is said synchronized with bottleneck μ_0 when $w_p(t) + \Delta_p^{\text{a}} = w_0(t)$. Bottlenecks μ_0 and μ_1 are always synchronized because they are accessible to every commuter.
- We note p^* the index of the most downstream bottleneck synchronized with bottleneck μ_p. From the moment when bottleneck μ_{p+1} starts to be used, it falls out of sync with bottlenecks μ_{p^*}, ..., μ_p when:
  - The flow of travelers coming from c_{p^*}, ..., c_p overcomes the sum of capacities μ_{p^*}, ..., μ_p, while the flow of travelers coming from collectors upstream of c_{p^*} is lower than the sum of capacities μ_{p^*}, ..., μ_{p+1}.

$$\begin{cases} \displaystyle\sum_{k=p^*}^{M} \dot{A}_k^{\text{th}}(t) \leq \sum_{k=p^*}^{p+1} \mu_k \\ \displaystyle\sum_{k=p^*}^{p} \dot{A}_k^{\text{th}}(t) > \sum_{k=p^*}^{p} \mu_k \end{cases} \tag{2.11}$$

In other words, the flow of travelers coming from c_{p^*}, ..., c_p cannot divert at bottlenecks strictly upstream of μ_p and make the additional travel time via bottlenecks μ_{p^*}, ..., μ_p increase higher than Δ^{a}_{p+1}. Meanwhile, the flow of travelers coming from collectors strictly upstream of c_p have bottleneck μ_{p+1} as an uncongested diversion option which additional travel time is lower than $w_p(t) + \Delta^{\text{a}}_p$. Consequently, bottleneck μ_{p+1} falls out of sync with the group μ_{p^*}, ..., μ_p. A **a**-div2 is observed on the group for the flows $\dot{A}^{\text{th}}_{p^*}(t)$, ... , $\dot{A}^{\text{th}}_p(t)$. A **a**-div1 is observed on bottleneck μ_{p+1} for the flows $\dot{A}^{\text{th}}_{p+1}(t)$, ..., $\dot{A}^{\text{th}}_M(t)$.

$$\begin{cases} \dot{A}_k(t) = \dfrac{\mu_k}{\sum_{l=p^*}^{p} \mu_l} \displaystyle\sum_{l=p^*}^{p} \dot{A}^{\text{th}}_l(t) & \forall k \in [\![p^*, p]\!] \\ \dot{A}_{p+1}(t) = \displaystyle\sum_{k=p+1}^{M} \dot{A}^{\text{th}}_k(t) \end{cases} \tag{2.12}$$

where $\dot{A}_{p+1}(t) \leq \mu_{p+1}$.

– Or the flow of travelers coming from collectors upstream of c_{p^*} overcomes the sum of capacities μ_{p^*}, ..., μ_{p+1}, while the ratio of the flow of travelers coming from c_{p^*}, ..., c_p on the sum of capacities μ_{p^*}, ..., μ_p overcomes the ratio of the flow of travelers coming from collectors strictly upstream c_p on μ_{p+1}.

$$\begin{cases} \displaystyle\sum_{k=p^*}^{M} \dot{A}^{\text{th}}_k(t) > \sum_{k=p^*}^{p+1} \mu_k \\ \dfrac{\sum_{k=p^*}^{p} \dot{A}^{\text{th}}_k(t)}{\sum_{k=p^*}^{p} \mu_k} > \dfrac{\sum_{k=p+1}^{M} \dot{A}^{\text{th}}_k(t)}{\mu_{p+1}} \end{cases} \tag{2.13}$$

In other words, the flow of travelers coming from c_{p^*}, ..., c_p cannot divert at bottlenecks strictly upstream of μ_p and make the additional travel time via bottlenecks μ_{p^*}, ..., μ_p increase at a higher pace than $w_{p+1}(t) + \Delta^{\text{a}}_{p+1}$. Meanwhile, the flow of travelers coming from collectors strictly upstream of μ_p have bottleneck μ_{p+1} as a congested diversion option. Consequently, bottleneck μ_{p+1} falls out of sync with the group μ_{p^*}, ..., μ_p. A **a**-div2 is observed on the group for the flows $\dot{A}^{\text{th}}_{p^*}(t)$, ... , $\dot{A}^{\text{th}}_p(t)$. A **a**-div2 is also observed on bottleneck μ_{p+1} for the flows $\dot{A}^{\text{th}}_{p+1}(t)$, ..., $\dot{A}^{\text{th}}_M(t)$. Travel times via bottlenecks of the group and μ_{p+1} are no longer equal due to two different waiting time increase paces. System 2.12 is still true but this time $\dot{A}_{p+1}(t) > \mu_{p+1}$.

- If none of conditions 2.11 and 2.13 are satisfied, bottleneck μ_{p+1} starts to be used in synchronization with the group.

  – If $\sum_{k=p^*}^{M} \dot{A}^{\text{th}}_k(t) \leq \sum_{k=p^*}^{p+1} \mu_k$, a **a**-div1 is observed on the incremented group of

bottlenecks (μ_{p^*}, ..., μ_{p+1}).

$$\begin{cases} \dot{A}_k(t) = \mu_k & \forall k \in [\![p^*, p]\!] \\ \dot{A}_{p+1}(t) = \sum_{k=p^*}^{M} \dot{A}_k^{\text{th}}(t) - \sum_{k=p^*}^{p} \mu_k \end{cases} \tag{2.14}$$

– If $\sum_{k=p^*}^{M} \dot{A}_k^{\text{th}}(t) > \sum_{k=p^*}^{p+1} \mu_k$, a **a**-div2 is observed on the incremented group of bottlenecks.

$$\begin{cases} \dot{A}_k(t) = \dfrac{\mu_k}{\sum_{k'=p^*}^{p+1} \mu_{k'}} \sum_{k'=p^*}^{M} \dot{A}_{k'}^{\text{th}}(t) & \forall k \in [\![p^*, p+1]\!] \end{cases} \tag{2.15}$$

- Re-synchronization between two groups of switched-on bottlenecks μ_{p^*}, ..., μ_p and μ_{q^*}, ..., μ_q ($p < q^*$) takes place when:

  – the demand from c_{p^*}, ..., c_p decreases to reach condition $\sum_{k=p^*}^{p} \dot{A}_k^{\text{th}}(t) < \sum_{k=p^*}^{p} \mu_k$. Then, the waiting times $w_{p^*}(t)$, ..., $w_p(t)$ decline until the complete re-synchronization of both groups: $w_p(t) + \Delta_p^{\text{a}} = ... = w_{p^*}(t) + \Delta_{p^*}^{\text{a}} = w_q(t) + \Delta_q^{\text{a}} = ... = w_{q^*}(t) + \Delta_{q^*}^{\text{a}}$.

  – the demand from c_{q^*}, ..., c_q increases or the demand from c_{p^*}, ..., c_p decreases so that $\frac{\sum_{k=p^*}^{p} \dot{A}_k^{\text{th}}(t)}{\sum_{k=p^*}^{p} \mu_k} < \frac{\sum_{k=q^*}^{q} \dot{A}_k^{\text{th}}(t)}{\sum_{k=q^*}^{q} \mu_k}$, i.e., the gap between additional travel times on the two groups of bottlenecks reduces until $w_p(t) + \Delta_p^{\text{a}} = w_{p^*}(t) + \Delta_{p^*}^{\text{a}} = ... = w_q(t) + \Delta_q^{\text{a}} = w_{q^*}(t) + \Delta_{q^*}^{\text{a}}$.

2.3.2.2 DUE with modes c, a and r available

In this second example, mode **r** is available on top of modes **c** and **a**. We describe the DUE on an example containing only two collectors and then generalize to any M value.

Example with M=2 We have $\Delta_k^{\text{a}} = \Delta_k^{\text{r}} + T_{\text{s}}$, $k \in \{1, 2\}$. We assume that $\Delta_1^{\text{r}} < \Delta_1^{\text{a}} < \Delta_2^{\text{r}} < \Delta_2^{\text{a}}$. Figure 2.5 presents the DUE described below.

1. Just as in the previous case, commuters initially choose **c** until $\dot{A}^{\text{th}}(t) > \mu_0$ and $w_0(t) = \Delta_1^{\text{r}}$.

2. As soon as $w_0(t) = \Delta_1^{\text{r}}$, travelers from c_1 start to use mode **r**. If $\dot{A}^{\text{th}}(t) - \dot{A}_1^{\text{th}}(t) \leq \mu_0$, a **r**-div1 occurs. A part of demand uses mode **c** while another part uses mode **r** (period A). We note $A_k^{\text{r}}(t)$ the cumulative count of travelers arriving/departing at/from c_k by mode **r**.

$$\begin{cases} \dot{A}_0(t) = \mu_0 \\ \dot{A}_1^{\text{r}}(t) = \dot{A}^{\text{th}}(t) - \mu_0 \end{cases} \tag{2.16}$$

3. When $\dot{A}^{\text{th}}(t) - \dot{A}_1^{\text{th}}(t) > \mu_0$, all travelers from c_1 choose mode **r**. This pattern is called **r**-full-div. Meanwhile, the waiting time on bottleneck μ_0 increases until $w_0(t) = \Delta_1^{\text{a}}$ (period B).

$$\begin{cases} \dot{A}_0(t) = \dot{A}^{\text{th}}(t) - \dot{A}_1^{\text{th}}(t) \\ \dot{A}_1^{\text{r}}(t) = \dot{A}_1^{\text{th}}(t) \end{cases} \tag{2.17}$$

4. As soon as $w_0(t) = \Delta_1^{\text{a}}$, it is worth for travelers from c_2 to use mode **a** with a transfer at c_1. If $\dot{A}^{\text{th}}(t) - \dot{A}_1^{\text{th}}(t) \leq \mu_0 + \mu_1$, we observe a **a**-div1 for the flow of travelers coming from c_2 only. Indeed, all commuters from c_1 keep using mode **r** which has a lower additional travel time (period C).

$$\begin{cases} \dot{A}_0(t) = \mu_0 \\ \dot{A}_1(t) = \dot{A}^{\text{th}}(t) - \dot{A}_1^{\text{th}}(t) - \mu_0 \\ \dot{A}_1^{\text{r}}(t) = \dot{A}_1^{\text{th}}(t) \end{cases} \tag{2.18}$$

5. As soon as $\dot{A}^{\text{th}}(t) - \dot{A}_1^{\text{th}}(t) > \mu_0 + \mu_1$, a **a**-div2 applies (period D) until $w_0(t) = w_1(t) + \Delta_1^{\text{a}} = \Delta_2^{\text{r}}$.

$$\begin{cases} \dot{A}_0(t) = \dfrac{\mu_0}{\mu_0 + \mu_1}(\dot{A}^{\text{th}}(t) - \dot{A}_1^{\text{th}}(t)) \\ \dot{A}_1(t) = \dfrac{\mu_1}{\mu_0 + \mu_1}(\dot{A}^{\text{th}}(t) - \dot{A}_1^{\text{th}}(t)) \\ \dot{A}_1^{\text{r}}(t) = \dot{A}_1^{\text{th}}(t) \end{cases} \tag{2.19}$$

6. A last **r**-div1 occurs for travelers coming from c_2 (period E) and lasts until demand decreases and queues vanish.

$$\begin{cases} \dot{A}_0(t) = \mu_0 \\ \dot{A}_1(t) = \mu_1 \\ \dot{A}_1^{\text{r}}(t) = \dot{A}_1^{\text{th}}(t) \\ \dot{A}_2^{\text{r}}(t) = \dot{A}^{\text{th}}(t) - \dot{A}_1^{\text{th}}(t) - \mu_0 - \mu_1 \end{cases} \tag{2.20}$$

Any M value Let us note p, the index of the most upstream collector at which mode **r** has a competitive additional travel time. Index p satisfies $\Delta_p^{\text{r}} \leq w_0(t) < \Delta_{p+1}^{\text{r}}$. Similarly, we note q the index of the most upstream collector at which mode **a** has a competitive additional travel time. Index q satisfies $\Delta_q^{\text{a}} \leq w_0(t) < \Delta_{q+1}^{\text{a}}$.

Because AMoD service time is necessarily positive, we have $\Delta_k^{\text{r}} < \Delta_k^{\text{a}}$ ($k \in [\![1, M]\!]$). Since mode **r** is uncongested and starts to be used by travelers from c_k before mode **a** with a transfer at c_k starts to be competitive, we do not find the same desynchronization effects as in the case with modes **c** and **a** only. Indeed, **r**-full-div pattern prevents travelers from c_k from crowding bottleneck μ_k and making the additional travel time via μ_k higher than via upstream bottlenecks. Consequently, all μ_k ($k \in [\![0, q]\!]$), are always synchronized here. The following rules characterize the DUE:

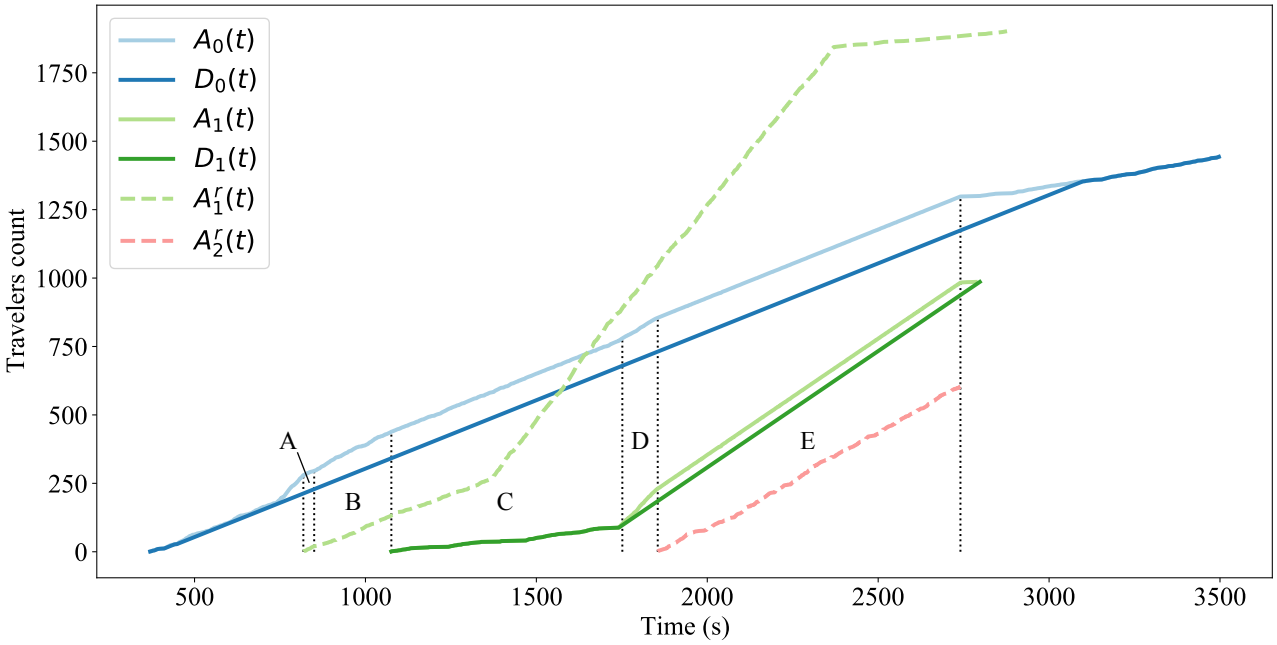


FIGURE 2.5 – *Arrival and departure curves presenting the DUE solution for the linear corridor model with $M = 2$, modes **c**, **r** and **a** available. Note that all the curves have been aligned in time on μ_0 to highlight the diversion patterns better. Diversion by mode **r** at c_1 starts from period A and it applies to all travelers departing from c_1 during following periods. Pattern **a**-div1 occurs on period C, E and **a**-div2 on period D.*

- As soon as $w_0(t) = w_1(t) + \Delta_1^{\mathrm{a}} = ... = w_q(t) + \Delta_q^{\mathrm{a}} = \Delta_{p+1}^{\mathrm{r}}$, the MRT station at c_{p+1} starts to be used.
  - If $\sum_{k=p+2}^{M} \dot{A}_k^{\mathrm{th}}(t) \leq \sum_{k=0}^{q} \mu_k$, a **r**-div1 pattern occurs. It is characterized by:

$$\begin{cases} \dot{A}_k(t) = \mu_k & \forall k \in [\![0, q]\!] \\ \dot{A}_k^{\mathrm{r}}(t) = \dot{A}_k^{\mathrm{th}}(t) & \forall k \in [\![1, p]\!] \\ \dot{A}_{p+1}^{\mathrm{r}}(t) = \dot{A}^{\mathrm{th}}(t) - \sum_{k=0}^{q} \mu_k - \sum_{k=1}^{p} \dot{A}_k^{\mathrm{th}}(t) & \end{cases} \tag{2.21}$$

  - If $\sum_{k=p+2}^{M} \dot{A}_k^{\mathrm{th}}(t) > \sum_{k=0}^{q} \mu_k$, a **r**-full-div pattern starts at c_{p+1}. It means that all travelers coming from c_{p+1} choose mode **r**. Meanwhile, **a**-div2 applies on the $q+1$ most downstream bottlenecks.

$$\begin{cases} \dot{A}_k(t) = \dfrac{\mu_k}{\sum_{k'=0}^{q} \mu_{k'}} \sum_{k'=p+2}^{M} \dot{A}_{k'}^{\mathrm{th}}(t) & \forall k \in [\![0, q]\!] \\ \dot{A}_k^{\mathrm{r}}(t) = \dot{A}_k^{\mathrm{th}}(t) & \forall k \in [\![1, p+1]\!] \end{cases} \tag{2.22}$$

- As soon as $w_0 = \Delta_{q+1}^{\mathrm{a}}$, bottleneck μ_{q+1} starts to be used for AV-to-MRT transfer.
  - If $\sum_{k=p+1}^{M} \dot{A}_k^{\mathrm{th}}(t) < \sum_{k=0}^{q+1} \mu_k$, a **a**-div1 pattern applies on the first $q+2$ bottlenecks

for the travelers coming from collectors strictly upstream of μ_p only.

$$\begin{cases} \dot{A}_k^{\mathrm{r}}(t) = \dot{A}_k^{\mathrm{th}}(t) & \forall k \in [\![1, p]\!] \\ \dot{A}_k(t) = \mu_k & \forall k \in [\![0, q]\!] \\ \dot{A}_{q+1}(t) = \sum_{k=p+1}^{M} \dot{A}_k^{\mathrm{th}}(t) - \sum_{k=0}^{q} \mu_k \end{cases} \tag{2.23}$$

– If $\sum_{k=p+1}^{M} \dot{A}_k^{\mathrm{th}}(t) \geq \sum_{k=0}^{q+1} \mu_k$, a **a**-div2 pattern applies on the first $q+2$ bottlenecks for the travelers coming from collectors strictly upstream of μ_p only.

$$\begin{cases} \dot{A}_k^{\mathrm{r}}(t) = \dot{A}_k^{\mathrm{th}}(t) & \forall k \in [\![1, p]\!] \\ \dot{A}_k(t) = \dfrac{\mu_k}{\sum_{k'=0}^{q+1} \mu_{k'}} \sum_{k'=p+1}^{M} \dot{A}_{k'}^{\mathrm{th}}(t) & \forall k \in [\![0, q+1]\!] \end{cases} \tag{2.24}$$

2.3.3 DUE properties

Note that the linear corridor model demonstrates desirable properties regarding the DUE. This section highlights the relationship between our network properties and the DUE properties.

First, our multimodal network fits the single destination "dividable" network definition of Tampere *et al.* (2010) proving the existence of DUE in such networks.

Second, the relaxed uniqueness of DUE[1] is ensured since our network has no more than one bottleneck per route (Iryo, 2013). Mounce (2009) proves that the solution set is a unique convex set based on the proof by Smith & Ghali (1990) that route cost functions are monotonous with respect to traffic volume in such networks. The strict monotonicity of route travel cost functions is required to claim strict uniqueness[2] of the DUE. Strictness prevents having two routes with equal travel costs that do not vary with respect to traffic flows variation. The latter situation produces several solutions for the DUE. Our network has characteristics that prevent falling into this case.

- The routes travel costs in free flow conditions are strictly ordered, i.e., the elements of the set $\{\Delta_k^{\mathrm{r}}, \Delta_k^{\mathrm{a}}\}_{k \in [\![1,M]\!]}$ are all different and non-null.
- When a route travel cost does not vary with respect to traffic flows variation (this is the case for r route, which has a constant travel cost, and for c and a routes when the bottleneck of the route works under capacity), it is necessarily compared with a route that has a strictly monotonous travel cost with respect to traffic flows variation at that time, i.e., that contains a congested bottleneck. We say that a bottleneck is congested when the slope of its departure curve equals its capacity (a queue exists or is just about to form at this bottleneck). In other words, among all routes sharing the same travel time, there is at most one route with a non-varying travel cost with respect to traffic flows variation. As a consequence, our system verifies the strictest definition of DUE uniqueness.

[1] DUE solution set is a unique convex set
[2] The incoming traffic flow profiles of all links are uniquely determined

Third, a robust solution methodology has been presented: it is simple and always converges.

2.4 Square corridor models

If the linear corridor model is convenient to define the DUE principles and describe the DUE with continuous flows of travelers, it does not account for how travelers access the collectors from their homes. In this section, we present the pseudo-dynamic and the dynamic square corridor models. They both take into account accessibility of collectors.

2.4.1 The pseudo-dynamic model

2.4.1.1 New assumptions

Figure 2.6 presents the assumptions made in the square corridor models and the associated parameters.

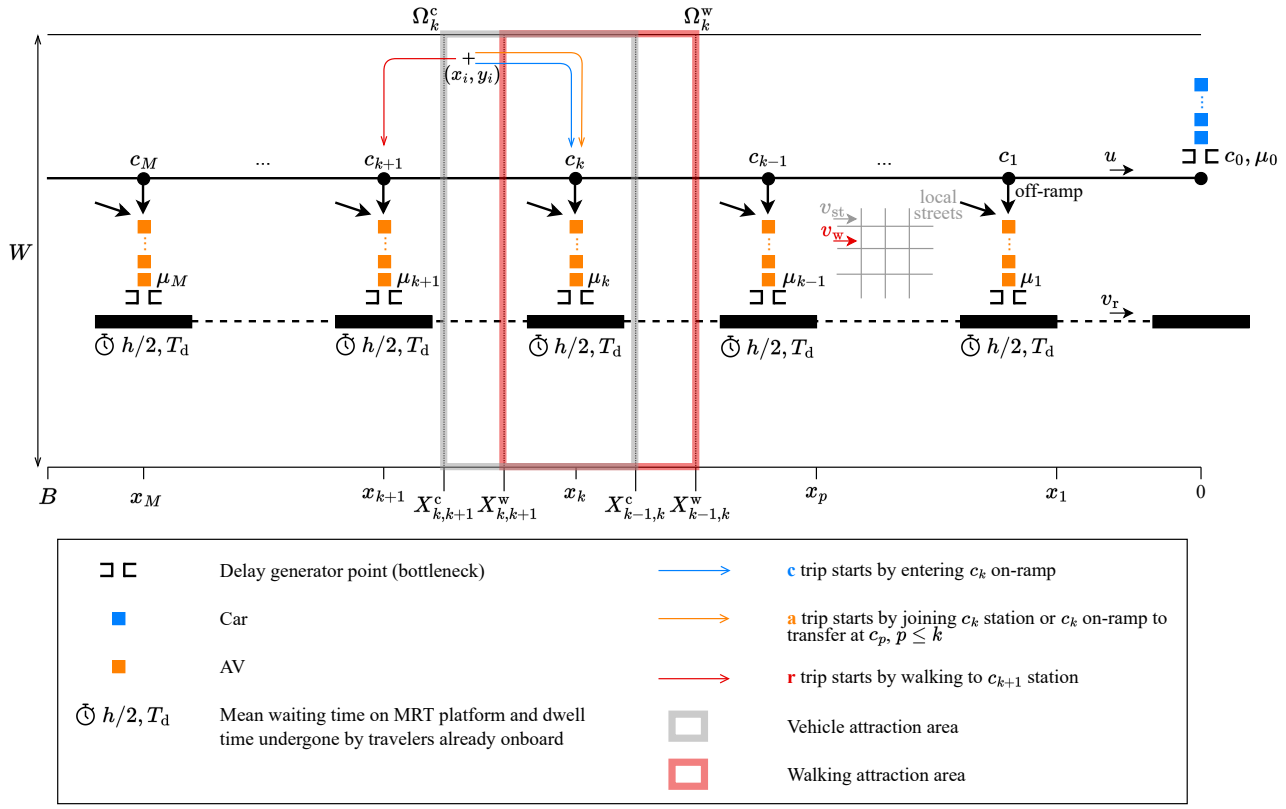


FIGURE 2.6 – *The square corridor models.*

Corridor geometry In the square corridor models, the corridor now has a width W along a y axis.

Demand spatial profile The origins of travelers are now distributed along the x and y axes. Traveler i's origin is located at (x_i, y_i). Collectors are connected to homes by an uncongested local streets grid that can be traveled at speed v_{st} by a vehicle and at speed v_{w} by walking. The three trip generation rates hold for the whole corridor area. The departure curves $D_{c_k}(t)$ are not defined directly as parameters of the square corridor models.

Transportation alternatives Now, four modes of transportation are available: car, MRT, AV, and walking. Three options, represented in Figure 2.7, are considered: car only (**c**), walking+MRT (**r**), and AV+MRT (**a**). As in the linear corridor model, each traveler chooses the mode and route that provides minimal travel time. Backward movements on the freeway are still prohibited. A traveler accessing the freeway by a vehicle at c_k, and accessing the MRT line by walking to $c_{k'}$ has to choose between $k+2$ itineraries:

- one corresponding to **c**: traveler drives until c_k, enters the freeway and keeps driving until c_0
- one corresponding to **r**: traveler walks to $c_{k'}$ station and takes the MRT until the terminus
- k corresponding to **a**: traveler rides an AV which joins c_k, then chooses to transfer to MRT at c_p, where $p \in [\![1, k]\!]$.

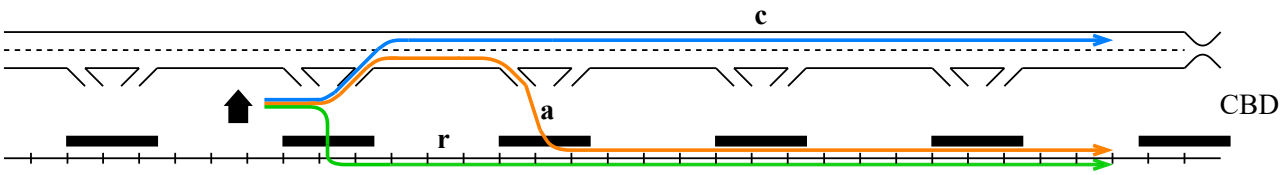


FIGURE 2.7 – *Transportation alternatives for the square corridor models are* **c**, **r**, **a** *respectively corresponding to car only, walking+MRT, AV+MRT trips.*

MRT operation The MRT operation is the same as in the linear corridor model.

AMoD operation For the pseudo-dynamic and the dynamic models, the AMoD service time experienced by traveler i is noted $T_{\text{s}}(t_i^{\text{req}})$, where t_i^{req} is the time at which i sends a request to AMoD. In the pseudo-dynamic model, the same assumption as in the linear corridor model is made on AMoD's service time: T_{s} is a constant function. In the dynamic model, T_{s} is no longer constant but time-variant and endogenously computed. Section 2.4.2.1 provides more details on the procedure to achieve this computation.

Delay generator points The same assumptions as in section 2.3.1.6 are made in the square corridor models. The non interaction of AVs and cars is now reasonable under two conditions. The first one is the same as in the linear corridor model. Second, cars and AVs should use different roads or lanes on the local network nearby MRT stations.

Access to infrastructure Each traveler has a fixed access collector by vehicle and a fixed access collector by walking. An access collector is chosen by the commuter to minimize their travel time in free-flow conditions.

The pool of commuters accessing the MRT line by walking to station k is called the walking attraction area of c_k and noted Ω_k^{w}. It forms a rectangle around c_k, gathering the origins of all commuters joining this collector rather than another by walking. Walking attraction areas boundaries can be computed by considering a traveler i departing from an origin located on the line defined as the intersection $\Omega_k^{\mathrm{w}} \cap \Omega_{k+1}^{\mathrm{w}}$ or as $x = X_{k,k+1}^{\mathrm{w}}$. On this line, travel times to c_k by accessing it directly or via c_{k+1} are equal.

Similarly, a vehicle attraction area rectangle Ω_k^{c} surrounds c_k. It gathers all commuters accessing the freeway via this collector by vehicle (car or AV) or directly accessing station k by AV. On the frontier $x = X_{k,k+1}^{\mathrm{c}}$ between Ω_k^{c} and $\Omega_{k+1}^{\mathrm{c}}$, travel times to c_k by entering freeway at on-ramp k or $k+1$ are equal for mode **c**. For mode **a**, with a given transfer collector c_p, travel times to c_p by accessing infrastructure via c_k or c_{k+1} are equal on the frontier $x = X_{k,k+1}^{\mathrm{c}}$.

Figure 2.6 provides an example of vehicle and walking attraction areas boundaries between c_k and c_{k+1} where $v_{\mathrm{w}} < v_{\mathrm{st}} < u$.

Travel times The travel times are now traveler-specific rather than collector-specific. The free-flow travel time by mode **c** for a commuter i departing from (x_i, y_i) is:

$$T_i^{\mathrm{c}} = \frac{|x_i - x_k| + |y_i|}{v_{\mathrm{st}}} + \frac{x_k}{u} \tag{2.25}$$

where c_k is the vehicle access collector of i.
The free-flow travel time by mode **a** when i transfers at c_p ($p \in [\![1, k]\!]$) is:

$$T_{i,p}^{\mathrm{a}} = \frac{|x_i - x_k| + |y_i|}{v_{\mathrm{st}}} + \frac{x_k - x_p}{u} + \frac{h}{2} + \frac{x_p}{v_{\mathrm{r}}} + (p-1)T_{\mathrm{d}} \tag{2.26}$$

Travel time by mode **r** is time-invariant:

$$T_i^{\mathrm{r}} = \frac{|x_i - x_{k'}| + |y_i|}{v_{\mathrm{w}}} + \frac{h}{2} + \frac{x_{k'}}{v_{\mathrm{r}}} + (k'-1)T_{\mathrm{d}} \tag{2.27}$$

where $c_{k'}$ is the walking access collector of i.

These free-flow travel time expressions allow to derive $X_{k,k+1}^{\mathrm{w}}$ and $X_{k,k+1}^{\mathrm{c}}$:

$$X_{k,k+1}^{\mathrm{w}} = \frac{1}{2}\left[x_{k+1} + x_k + (x_{k+1} - x_k)\,\frac{v_{\mathrm{w}}}{v_{\mathrm{r}}} + T_{\mathrm{d}} v_{\mathrm{w}}\right] \tag{2.28}$$

$$X_{k,k+1}^{\mathrm{c}} = \frac{1}{2}\left[x_{k+1} + x_k + (x_{k+1} - x_k)\,\frac{v_{\mathrm{st}}}{u}\right] \tag{2.29}$$

where boundaries are well defined between two collectors when the spacing s_k between c_k and c_{k+1} verifies $s_k > T_{\mathrm{d}} v_{\mathrm{w}} v_{\mathrm{r}} / (v_{\mathrm{w}} - v_{\mathrm{r}})$.

Outside of free-flow conditions, the travel time of i by mode **c** is:

$$\tau_i^{\mathrm{c}} = T_i^{\mathrm{c}} + w_0(t_i^0) \tag{2.30}$$

where t_i^0 is the theoretical arrival time of i at c_0 (when i travels by car in free-flow conditions). Outside of free-flow conditions, the travel time by mode **a** when i transfers at c_p ($p \in [\![1, k]\!]$) is:

$$\tau_{i,p}^{\mathrm{a}} = T_{\mathrm{s}}(t_i^{\mathrm{req}}) + T_{i,p}^{\mathrm{a}} + w_p(t_i^p) \tag{2.31}$$

where t_i^p is the arrival time of i at c_p.

Moreover, free-flow additional travel times by modes **r** and **a** compared to mode **c** are $\Delta_i^{\mathrm{r}} = T_i^{\mathrm{r}} - T_i^{\mathrm{c}}$, and $\Delta_{i,p}^{\mathrm{a}} = T_{\mathrm{s}}(t_i^{\mathrm{req}}) + T_{i,p}^{\mathrm{a}} - T_i^{\mathrm{c}}$ respectively.

Equilibrium The same assumptions as in section 2.3.1.8 are made in the square corridor models.

2.4.1.2 Dynamic user equilibrium solution

In this section, we present and justify the DUE numerical resolution process used by the pseudo-dynamic model.

FIFO property Because all travelers undergo the same AMoD service time and have a defined vehicle access collector, the FIFO rule holds in the system. If traveler i arrives before traveler j at bottleneck μ_0 in free-flow conditions ($T_i^{\mathrm{c}} < T_j^{\mathrm{c}}$), then i also arrives before j at bottleneck μ_p where $p \leq \min(k, k')$, c_k and $c_{k'}$ being the vehicle access collector of i and j respectively. The order of arrivals at bottleneck μ_0 is conserved at bottlenecks μ_p, $p \in [\![1, M]\!]$. The route choice of a traveler depends only on the choices made by travelers that have arrived before them at bottleneck μ_0 in free-flow conditions.

Resolution process The FIFO property allows the direct exact computation of the equilibrium by processing travelers' itinerary choices in the order of free-flow arrivals at the CBD off-ramp. This resolution process is said to be numerical because it no longer considers travelers aggregated in flows but travelers one by one to find the Nash equilibrium solution. The Nash equilibrium is the counterpart of Wardrop's equilibrium, considering travelers individually. This resolution process always finds a solution: it is robust.

DUE principles Figure 2.8 presents the DUE solution for the pseudo-dynamic model with two collectors and a uniform distribution of travelers along the axes. One can observe the same kinds of diversion patterns as in the linear corridor model.

Indeed, if $D_{c_k}(t)$ is not a parameter in the square corridor models, one can build $D_{c_k}^{\mathrm{th}}(t)$ by assuming that all travelers departing from Ω_k^{c} choose mode **c**. Then, $A_k^{\mathrm{th}}(t)$ ($k \in [\![1, M]\!]$) and $A^{\mathrm{th}}(t)$ are obtained the same way as in the linear corridor model.

As in the linear corridor model, $\Delta^{\mathrm{a}}_{i,p}$, does not depend on i in the pseudo-dynamic model because the free-flow in-vehicle access times by modes **c** and **a** are equal and T_{s} is a constant function. Consequently, the sequence of **a** diversion patterns is similar: the system also alternates between **a**-div1 and **a**-div2. Since we are still treating a many-to-one demand pattern, bottlenecks can become out of sync.

One difference with the linear corridor model is that the additional travel time by mode **r** is traveler specific. When travel times via downstream bottlenecks reach Δ^{r}_i, mode **r** starts to be worthwhile for traveler i. They choose r as long as travel times via downstream bottlenecks are strictly greater than Δ^{r}_i. It is an equivalent of the **r**-full-div pattern but for traveler i only. As the system loads, a rhombus surrounding each collector widens, gathering more and more travelers preferring mode **r** over the other options.

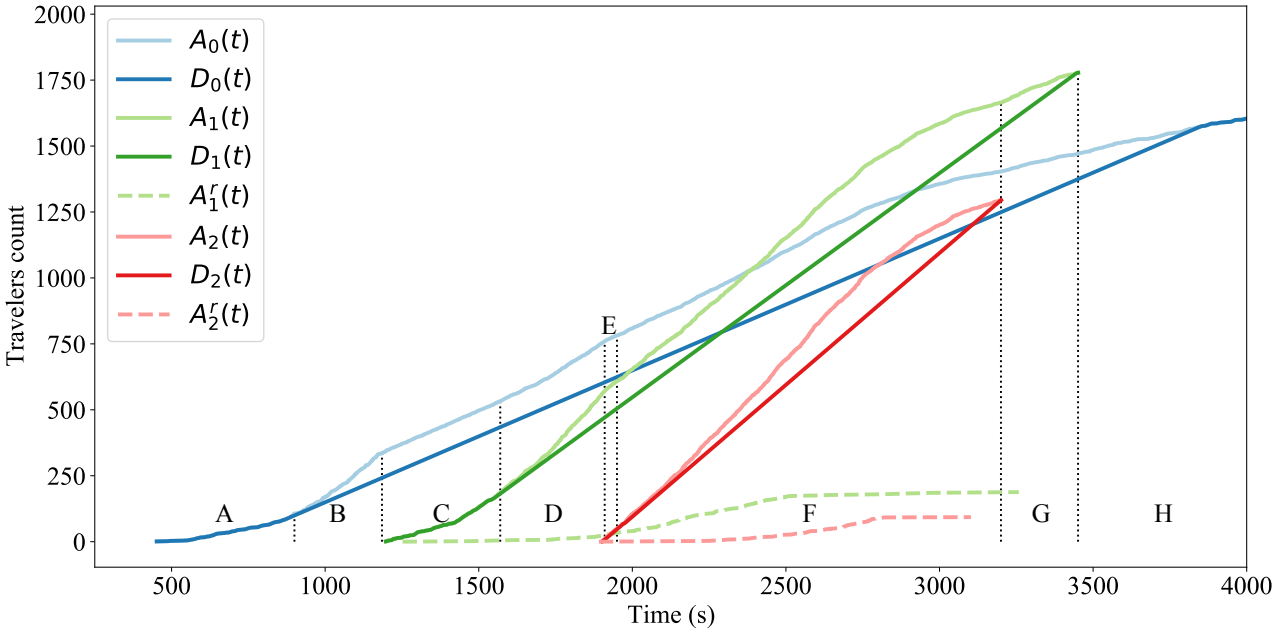


FIGURE 2.8 – *Arrival and departure curves presenting the DUE solution for the pseudo-dynamic model in a corridor where M=2. Note that all the curves have been aligned in time on c_0 to highlight the diversion patterns better. Congestion on bottleneck μ_0 starts at the beginning of period B. In period C, a **a**-div1 pattern occurs with c_1 as a transfer collector. Pattern **a**-div2 occurs in period D when bottleneck μ_1 becomes congested. The state shifts continue in periods E and F until the decrease in demand leads to the desertion of c_2 in period G and c_1 in period H. Diversion to **r** progressively concerns more travelers as waiting times at bottlenecks rise.*

2.4.2 The dynamic model

The constant service time assumption is acceptable when the fleet size is large enough. In general, AMoD service time depends on dispatching rules, the number of vehicles m in the fleet, and the amount of work to achieve, i.e., the distance to serve customers and relocate empty vehicles. The dynamic model goes one step further than the pseudo-dynamic model by relaxing the constant service time assumption. The other assumptions regarding corridor geometry, demand spatial profile, transportation alternatives, MRT operation, delay generator points, access to infrastructure and equilibrium are the same as in the pseudo-dynamic

model. In the dynamic model, DUE resolution and service time computation are nested in an iterative process.

2.4.2.1 New assumptions and resolution process

AMoD dispatching strategy A dispatching heuristic traditionally used in (A)MoD systems is the *nearest-idle-vehcile* (Maciejewski *et al.*, 2016). It conforms to the FIFO rule by treating requests in the order of arrival. The first traveler that has ordered an AV is the first to be assigned a vehicle by the dispatcher. We consider such a dispatching strategy here. As no booking is allowed, $t_i^{\text{req}} = t_i^{\text{dep}}$ for each traveler i choosing **a**, where t_i^{dep} is the time at which traveler i departs from home. The first traveler to depart is also the first to send a request. Service time is a function of the request/departure time.

MSA process To approximate the service time profile and compute the DUE, we solve the fixed point problem $\widehat{T}_{\text{s}}(\text{DUE}(T_{\text{s}})) = T_{\text{s}}$ with the Method of Successive Averages (MSA) as detailed in algorithm 1.

In step 1 of the algorithm, an initial service time profile should be chosen. It can be defined arbitrarily or computed in a prior iteration. During this prior iteration, the service time is computed online by considering that AMoD receives requests in the order of theoretical arrival at c_0. We make the following strong assumption: $t_i^{\text{req}} = f(t_i^0)$ where f is a linear function. Each time a traveler chooses a route, the service time is re-computed following equation 2.32. Service time is updated only if the difference between the new and last values is above a certain threshold. It forms a step function. Once all travelers have chosen a route, they are sorted in the order of departure time, which is the effective order in which AMoD receives requests. As route choices are unchanged, the effective service time profile $\widehat{T}_{\text{s}}$ is computed considering this order and following equation 2.32. It serves as the reference initial service time profile (or predicted service time profile) for the subsequent iteration of MSA ($T_{\text{s}} \leftarrow \widehat{T}_{\text{s}}$ at the end of the prior iteration).

Step 3 of the algorithm launches the MSA loop which contains at least one iteration. The quality criteria used as stopping conditions for the loop contain conditions on the difference between the predicted and the effective service time profiles, and on the quality of the DUE solution found. They are detailed in section 2.4.2.2.

In step 5, the predicted service time profile is a parameter for DUE resolution. Each traveler chooses an itinerary knowing the service time they will experience on mode **a**. The same numerical resolution process as in the pseudo-dynamic model is used here. Section 2.4.2.2 discusses the relevance of applying such a process in a system where FIFO rule is broken due to the time-variant AMoD service time.

In step 6, the effective service time profile $\widehat{T}_{\text{s}}$ is computed the same way as in the prior iteration, considering the distribution of travelers on itineraries resulting from step 5.

In step 7, the predicted service time for iteration $K+1$ of MSA is computed based on predicted and effective service time profiles of iteration K.

Service time computation Equation 2.32 is used to compute the reference initial service time profile in the prior iteration and the effective service time profile in step 6 of the algorithm.

$$T_s(t_j^{req}) = 2E_2 + E_3 - \left[t_i^{req} + \frac{\Delta t}{m} - E_1 \right] \tag{2.32}$$

It expresses the service time of traveler j as a function of E_1, E_2, E_3 and traveler i's parameters, i and j being processed consecutively. It is an approximation based on the last m requests $\{l\}$ received by the fleet during Δt. If i is the last traveler processed, $\Delta t = t_i^{req} - \min_l(t_l^{req})$. As AMoD treats requests in their order of reception, the next traveler choosing this fleet will necessarily ride one of the AVs serving $\{l\}$, a mean predecessor request, representative of $\{l\}$, is built. With $E[.]$ being the expected value, the mean request is picked up at $E_1 = E[t_l^{req} + T_s(t_l^{req})]$, rides for $E_2 = E[\frac{|x_l - x_{k(l)}| + |y_l|}{v_{st}} + \frac{x_{k(l)} - x_{p(l)}}{u}]$ and waits for drop-off for $E_3 = E[w_{p(l)}]$ where $k(l)$ and $p(l)$ respectively label vehicle access collector and transfer collector for traveler l. AV relocates for E_2 to finally serve the next request, which is expected to be received by AMoD at $t_i^{req} + \frac{\Delta t}{m}$. AVs are initially located at the m first requests locations so that the service time of a fleet remains null until the mth request is received.

Algorithm 1: MSA process

1 Initialize T_s with an arbitrary step function or with a prior iteration (reference initial service time profile);
2 $K \leftarrow 1$;
3 **while** $K = 1$ *or DUE(T_s) does not meet the quality criteria* **do**
4 $\quad K \leftarrow K + 1$;
5 $\quad$ Compute DUE(T_s) by processing travelers one by one, taken in order of increasing t_i^0;
6 $\quad$ Compute $\widehat{T_s}$(DUE(T_s)) based on equation 2.32 by processing travelers by increasing $t_i^{req} = t_i^{dep}$;
7 $\quad T_s \leftarrow \frac{1}{K}\widehat{T_s} + (1 - \frac{1}{K})T_s$;
8 Return DUE(T_s);

2.4.2.2 DUE properties

With a time-variant endogenous service time, the formal demonstration of the existence and uniqueness of the DUE is challenging.

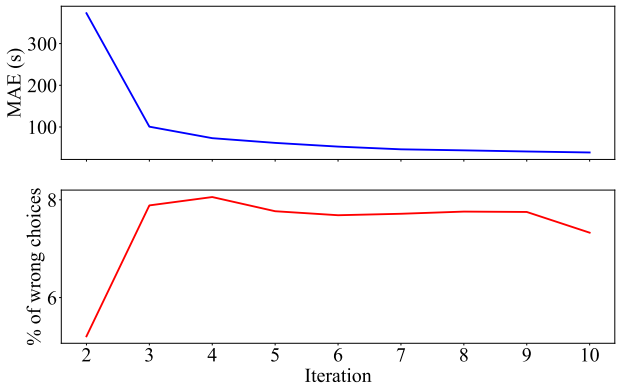
Non-FIFO As service time is a continuous function of departure time, it is a discontinuous function of theoretical arrival time at bottleneck μ_0. The service time profile impacts the waiting times at drop-off bottlenecks w_k ($k \in [\![1, M]\!]$). When service time increases between two travelers i and j taken in the order of theoretical arrival at bottleneck μ_0 ($t_i^0 < t_j^0$), this order is maintained at drop-off bottlenecks and the predicted **a** travel time equals those realized. When service time decreases, travelers may be re-ordered: a customer requesting an AV later may arrive earlier at the transfer collector. A few commuters choose a non-optimal route because their waiting times at the AV-to-MRT transfer are longer than expected due to order rearranging.

Existence and uniqueness Discontinuities in the service time profile are local and bounded by the corridor parameters: the service time globally follows the loading of the network. It gives us confidence that jumps in service time have limited impact compared to the periods when T_s is constant and that the existence and uniqueness of the DUE still hold. However, we were not able to elaborate a formal proof.

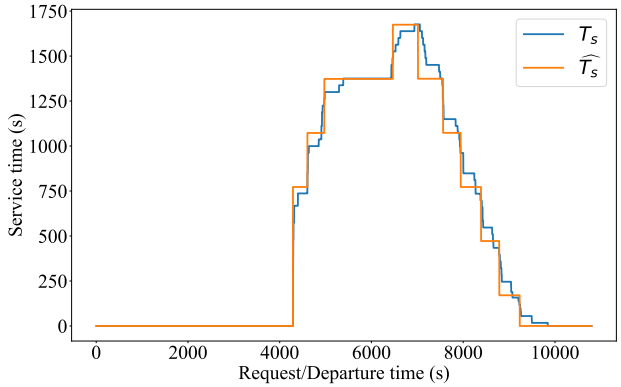
Quality criteria Numerically, three criteria assess the convergence of the process and the quality of the solution found. The quality criteria (step 3) are the mean absolute error (MAE) between $\widehat{T}_s$ and T_s, quartiles of the difference $\widehat{T}_s - T_s$, and the percentage of travelers that have made a wrong route choice due to local order rearranging. The process found a solution for each experiment achieved with the following criteria: MAE below 40s, 1st and 3rd quartiles respectively longer than -5min and shorter than 5min, less than 10% wrong choices.

Numerical experiments Figures 2.9 and 2.10 present one of the tests for a generic scenario. The scenario deals with a theoretical monocentric city where a high-frequency subway serves the whole corridor. The demand level is high, with 15.5k travelers in 3 hours, leading to intense congestion from the second hour. The discontinuity threshold on $\widehat{T}_s$ is 5min. Table 2.3 provides the scenario parameters.

Replacing the reference initial conditions (computed during the prior iteration of MSA as described above) with different initial service time profiles leads to very close equilibrium. Figure 2.10b compares the distribution of travelers obtained after convergence for the reference initial service time profile and a constant profile. Figure 2.10a shows the initial and final service time profiles. The MAE between final profiles is 57s. The maximum variation of the volume of travelers on a route represents only 0.28% of the total volume. The other initial service time profiles tested (translated reference, null, one-step function, three-steps function) led to similar results.

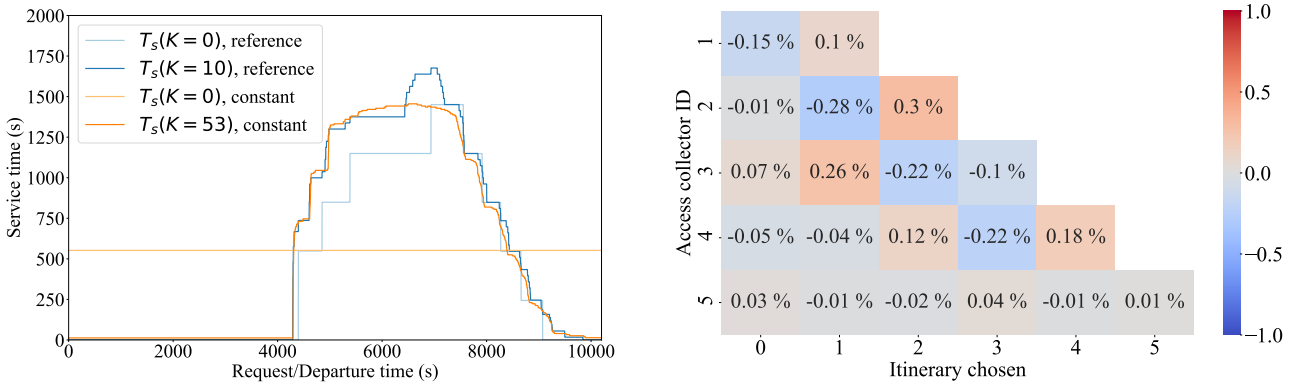


(a) *Evolution of MAE($\widehat{T}_s - T_s$) and percentage of travelers that have made a wrong route choice.*



(b) *Difference between $\widehat{T}_s$ and T_s is acceptable at iteration 10 of the MSA process.*

FIGURE 2.9 – *Convergence of the endogenous time-variant AMoD service time with the initial reference conditions.*



(a) *Initial and final service time profiles.*

(b) *The relative differences in the travelers' distribution are marginal.*

FIGURE 2.10 – *Comparison of equilibrium obtained with reference and constant initial service time profiles.*

The number of iterations required to meet the quality criteria differs depending on the initial conditions. The prior iteration used to compute the reference initial conditions provides a good starting point. The process reaches the quality criteria in only ten iterations, as shown in Figure 2.9a. This fast convergence enhances compatibility with the design optimization framework.

Scenario name	Monocentric city
Description	Wide monocentric compact city with its suburbs
Distribution of travelers	Uniform
MRT type	High frequency subway
B * W **(km^2)**	20*3
Morning peak duration (h)	3
Number of travelers	15.5k
v_{w} **(m/s)**	1.2
v_{st} **(m/s)**	9
u **(m/s)**	19
v_{r} **(m/s)**	14
h **(s)**	240
T_{d} **(s)**	45
μ_k**,** $0 < k \leq M$ **(veh/s)**	0.2
μ_0 **(veh/s)**	0.6
m **(veh)**	1550
Number of collectors	5
s_k**,** $0 \leq k < M$ **(km)**	4
T_{s} **step size (s)**	300

TABLE 2.3 – *Parameters for the monocentric city scenario.*

2.5 A static model for benchmarking

Traditionally, transportation studies use static approaches with time-invariant supplies and demand flows to deal with the morning commute problem in a corridor (Vuchic & Newell, 1968, Vuchic, 1969, Wirasinghe *et al.*, 1977). In this section, we formulate a static model to treat the square corridor case. This model is the static counterpart of the dynamic model. It is used in the next chapter for benchmarking to show how crucial considering dynamics is.

Commuters aggregation In the static model, the square corridor splits into N equally proportioned zones aggregating the origins of travelers in an abstract origin O_i. The access travel time from O_i to c_k by vehicle (and walking respectively) is defined as the average vehicle (and walking respectively) access time of aggregated commuters. O_i's vehicle (and walking respectively) access collector corresponds to the one with the minimal vehicle (and walking respectively) access time.

Multimodal network directed graph The corridor is equivalent to a directed graph $G = (V, E)$ where V is the set of vertices and E the set of edges. Vertices include origins $\{O_i\}_{i\in[\![1,N]\!]}$, destination $\{D\}$ and three different nodes for each collector $\{C_k^{\mathrm{c}}, C_k^{\mathrm{r}}, C_k^{\mathrm{a}}\}_{k\in[\![1,M]\!]}$. Commuters whose vehicle (and walking respectively) access collector is c_k use C_k^{c} (and C_k^{r} respectively) when they choose mode **c** (and mode **r** respectively). Commuters choosing mode **a** with a transfer at c_k use C_k^{a} then C_k^{r}. The edges of the graph link:

- origins to their vehicle and walking access collectors,
- origins to transfer nodes associated with collectors located downstream of their vehicle access collector,
- **a** nodes of collectors to their **r** nodes,
- **c** nodes of collectors to C_1^{c},
- **r** nodes of collectors to destination,
- C_1^{c} to destination.

Figure 2.11 represents an example of G.

Travel times Travel times are constant except for edges passing through a delay generator point, namely (C_1^{r}, D) and $(C_k^{\mathrm{a}}, C_k^{\mathrm{r}})$, $k \in [\![1, M]\!]$. The BPR volume delay function (Maerivoet & De Moor, 2005) gives the travel times on these links. Two sets of parameters $(\alpha_{\mathrm{c}}, \beta_{\mathrm{c}})$ and $(\alpha_{\mathrm{a}}, \beta_{\mathrm{a}})$ determine the shape of the BPR functions for both delay generator point types (freeway off-ramp to the CBD and AV-to-MRT transfer). Volumes and travel times are noted $\nu(\text{edge})$ and $\tau(\text{edge})$ respectively. AMoD service time T_{s} is constant in the static model.

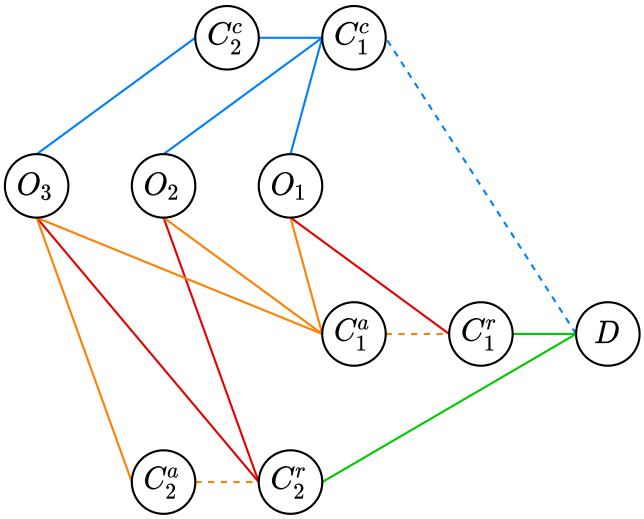


FIGURE 2.11 – *Graph G representing the corridor in the static model. In this example, N=3, M=2, c_1 is the vehicle access collector of O_1 and O_2, the walking access collector of O_1, c_2 is the vehicle access collector of O_3, the walking access collector of O_2 and O_3. Solid edges carry constant travel times, while volume-dependent travel times are associated with dashed edges. Colors represent the mode used to travel the edge (car in blue, AV in orange, MRT in green, and walking in red).*

$$\tau(O_i, C_k^{\mathrm{c}}) = \sum_{j \in O_i} \frac{|x_j - x_k| + |y_j|}{\mathrm{card}(O_i) v_{\mathrm{st}}} \tag{2.33}$$

$$\tau(O_i, C_k^{\mathrm{r}}) = \sum_{j \in O_i} \frac{|x_j - x_k| + |y_j|}{\mathrm{card}(O_i) v_{\mathrm{w}}} \tag{2.34}$$

$$\tau(O_i, C_p^{\mathrm{a}}) = T_{\mathrm{s}} + \sum_{j \in O_i} \frac{|x_j - x_k| + |y_j|}{\mathrm{card}(O_i) v_{\mathrm{st}}} + \frac{x_k - x_p}{u} \tag{2.35}$$

$$\tau(C_k^{\mathrm{a}}, C_k^{\mathrm{r}}) = T_{\mathrm{f}} \left(1 + \alpha_{\mathrm{a}} \left(\frac{\nu(C_k^{\mathrm{a}}, C_k^{\mathrm{r}})}{\mu_k} \right)^{\beta_{\mathrm{a}}} \right) \tag{2.36}$$

$$\tau(C_k^{\mathrm{r}}, D) = \frac{h}{2} + \frac{x_k}{v_{\mathrm{r}}} + (k-1) T_{\mathrm{d}} \tag{2.37}$$

$$\tau(C_k^{\mathrm{c}}, C_1^{\mathrm{c}}) = \frac{x_k - x_1}{u} \tag{2.38}$$

$$\tau(C_1^{\mathrm{c}}, D) = \frac{x_1}{u} + \alpha_{\mathrm{c}} \left(\frac{\nu(C_1^{\mathrm{c}}, D)}{\mu_0} \right)^{\beta_{\mathrm{c}}} \tag{2.39}$$

Solving process The equilibrium of this system is computed with an MSA using the conventional descent gradient based on an all-or-nothing assignment and step size $\frac{1}{K}$ where K is the iteration number of the method (Sheffi & Powell, 1982).

Static model calibration Parameters of the static model should be calibrated on the dynamic model outputs to maintain a certain consistency and facilitate the comparison of the

results. N is chosen to keep pools of travelers small enough to account for access times diversity (necessarily smaller than attraction areas) and wide enough to gather a sufficient number of commuters for the flow balance procedure. T_s value is set to the mean service time computed by the dynamic model on the same scenario. Note that the congestion term in $\tau(C_1^c, D)$ (2.39) is intentionally independent of the free-flow travel time on edge (C_1^c, D), which is controlled by the design parameter x_1. It prevents the unwanted modification of the congestion function when the corridor design changes, as it does in the next chapter. The free-flow travel time on (C_1^c, D) under the nominal design is used for α_c. The other parameters (α_a, β_a, β_c) are chosen to provide consistent total travel time and waiting times at bottlenecks regarding the dynamic model on the same scenario.

2.6 Conclusion and discussion

In this chapter, we introduced four versions of a model to study the morning commute in a corridor (Table 2.4). The first version presented, the linear corridor model, considers flows of travelers departing from the collectors and assumes a fixed AMoD service time for all. It allows focusing on the bottlenecks equilibrium states without considering the accessibility issues to join the freeway and the MRT line. We described the DUE precisely in two cases, one with mode **r** available and another without mode **r**. The concepts of synchronization/desynchronization of bottlenecks and four diversion patterns (**a**-div1, **a**-div2, **r**-div1, and **r**-full-div) were graphically illustrated on examples with only two collectors and analytically defined for any number of collectors.

The second and third versions presented, the square corridor models, account for accessibility issues by distributing travelers' origins within a square corridor. A simple numerical resolution process provides the exact DUE in the pseudo-dynamic model, which considers a constant AMoD service time profile. A fast iterative resolution process provides a good quality solution to the DUE in the dynamic mode. This process endogenously approximates the time-variant AMoD service time profile.

Finally, the fourth version is a static model that does not take into account the dynamics of the system. In the next chapter, the pseudo-dynamic, dynamic and static models are compared to study the impact of the corridor design on the cooperation or competition between MRT and AMoD.

The models presented include several strong assumptions and are quite restrictive regarding intermodality and network. However, the simplifications performed make the DUE dynamics explicit and allow getting round the black-box effect that characterizes less restrictive frameworks. The network structure is generic enough for application to several urban areas. Our models could be extended to more intermodal options. Extensions should maintain the network properties (single destination, single bottleneck per route) and the ability of our UE resolution procedure to meet the quality criteria in a few iterations. They should either preserve the theoretical order of arrival at all bottlenecks (as in the linear corridor and pseudo-dynamic models) or generate local bounded order rearranging (as in the dynamic model). For example, a car + MRT option with capacitated park-and-rides at collectors can be added. An active mobility + MRT option would be similar to mode **r**.

Model	**Morning commute dynamics**	**Travelers disaggregation**	**Endogenous time-variant AMoD service time**	**DUE resolution**
Static	✗	✗	✗	Approximated - Travelers flows balance in a graph with MSA
Linear corridor	✓	✗	✗	Exact - Direct application of Wardrop's first principle on flows of travelers
Pseudo-dynamic	✓	✓	✗	Exact - Sequential individual choices of travelers in the order of theoretical arrivals at c_0 in free-flow conditions
Dynamic	✓	✓	✓	Approximated - Integrated AMoD service time profile computation with MSA and DUE resolution with sequential individual choices of travelers

TABLE 2.4 – *Four versions of the model to study the morning commute in a corridor.*

3. THE JOINT MRT-AMOD DESIGN PROBLEM

The elements of this chapter have been presented in:

- Cortina, M., Chiabaut, N., Leclercq, L. (2023). Fostering synergy between transit and Autonomous Mobility-on-Demand systems: A dynamic modeling approach for the morning commute problem. *Transportation Research Part A: Policy and Practice*, 170:103638. https://doi.org/10.1016/j.tra.2023.103638.

3.1 Motivation and objectives

In chapter 2, we introduced four versions of a model to study the morning commute in a multimodal corridor. They all account for the impact of congestion on the distribution of travelers over the different modes and routes but do not consider the system's dynamics equally. The static model does not consider demand variations in time, congestion dynamics, and ignores the coupling between demand and AMoD supply, contrary to the dynamic model. The first objective of this chapter is to compare the static and dynamic models in terms of their sensitivity to the corridor design, especially to the number of collectors with constant spacing. With the sensitivity analysis of the number of collectors on the volume-based and distance-based mode share, we aim to highlight the importance of taking into account the dynamics for design purposes and understand better the circumstances of cooperation and competition between AMoD and MRT.

The linear corridor model allowed us to analytically and graphically illustrate the DUE principles. As highlighted in section 2.4.1.2 of chapter 2, the pseudo-dynamic model presents the same diversion patterns and desynchronization effects as the linear corridor model. The direct application of the DUE principles to the dynamic model is not possible because the system no longer conforms to the FIFO rule. Hence, the second objective of this chapter is to analyze the impact of the corridor design on cooperation and competition between AMoD and MRT with the pseudo-dynamic model in the light of DUE principles. Moreover, we compare the pseudo-dynamic and dynamic models.

Based on these analyses, we use the dynamic model to address the question of the joint design of the MRT and AMoD. Several decision variables have been investigated in the PT-(A)MoD design literature, as shown in Figure 3.1. On the PT side, the choice of the legs to

operate and the frequency or headway of PT lines are typical decision variables. The number of MRT stations, the spacing between stations, and the MRT line alignment have been considered in corridors. On the AMoD side, the fleet size is the main decision variable optimized. Only Liu & Ouyang (2021) consider the size of a square coverage zone as a degree of freedom. In this chapter, we choose traditional MRT design parameters, namely the number of stations and their locations, in addition to AMoD design parameters related to the geofencing regulation measure, namely the number of fleets of AVs and the boundaries of their coverage zones. To our knowledge, this combination of decision variables has not been investigated yet in the literature. It is particularly adapted to our corridor case study. So the third objective of this chapter is to propose a framework to optimize this combination of decision variables in a realistic scenario. We compare three policies in which the transportation authority (TA) forbids AMoD, authorizes AMoD under regulation, and authorizes AMoD without any regulation. In this way, we address the fourth objective of this chapter which is to evaluate the benefits of a multimodal system based on fostered cooperation between the MRT and AMoD.

	PT/MRT	AMoD
THEORETICAL NETWORK	- Number of MRT stations - Spacing of MRT stations - Alignment of the MRT line - Headway/Frequency of the MRT line	- Fleet size - Size of a square coverage zone
REAL NETWORK	- PT legs to operate - Frequency of each PT line	- Fleet size

FIGURE 3.1 – *Typical decision variables in the PT-(A)MoD design literature.*

3.2 Notations for this chapter

Table 3.1 only gathers the notations introduced in the present chapter. All notations introduced in chapter 2 maintain here.

TABLE 3.1 – *Notations for chapter 3.*

Notation	Definition
Λ	Total distance traveled by all modes in the corridor
$\xi_i^{\mathrm{c,a}}$	Binary variable which equals 1 if traveler i has chosen mode **c** or **a**, and 0 otherwise
ξ_i^{r}	Binary variable which equals 1 if traveler i has chosen mode **r**, and 0 otherwise

Continue on the next page

TABLE 3.1 – *Notations for chapter 3 (continued).*

Notation	Definition
Λ_x, Λ_y	Components of Λ corresponding to the total distances traveled along x and y axis respectively
Ω_k^{c-}	Part of Ω_k^c located downstream x_k
Ω_k^{c+}	Part of Ω_k^c located upstream x_k
D	Sum of maximal detours for all attraction areas of the corridor, where a detour refers to a backward movement along x axis
Ω^*	Union of all intersections between the vehicle attraction area of a collector and the walking attraction area of its neighboring upstream collector

3.3 Comparison of static and dynamic models in terms of sensitivity to the MRT design

This section compares models in terms of their sensitivity to the corridor design. The sensitivity analysis is conducted on the theoretical monocentric city scenario, which parameters for the dynamic model have been provided in the previous chapter by Table 2.3. Table 3.2 provides the parameters of the same scenario for the static model, calibrated on the criteria described in section 2.5.

Scenario name	Monocentric city
α_c	222
β_c	4
α_a	1
β_a	2
Exogenous constant T_s (s)	920
Pool dimensions (m^2)	300*300

TABLE 3.2 – *Parameters of the monocentric city scenario for the static model.*

A crucial design parameter in our case study is the number of collectors and their locations along the corridor. An *a priori* fair design strategy regarding a uniform distribution of travelers, as in the monocentric city scenario, is to fix the spacing between two collectors. We analyze the sensitivity of the number of collectors M with constant spacing on system-level and link-level indicators. The system-level indicators are volume-based and distance-based modal shares. A distance-based mode share corresponds to the total distance traveled by this mode over the total distance traveled by all modes. The link level indicators are the flows on each mode-route alternative.

Note that the TA does not regulate AMoD here. As presented in chapter 2, AMoD comprises a single fleet operating on the whole corridor.

3.3.1 Effect of the number of collectors on the total distance traveled

Before comparing models on distance ratios, one can verify if the investigated design parameter has the same effect on the total distance traveled in both cases.

Let us note Λ the total distance traveled by all modes in the corridor. It can be expressed as follows:

$$\Lambda = \sum_{k=1}^{M} \left(\sum_{i \in \Omega_k^c} \left(\xi_i^{c,a} (|x_i - x_k| + x_k) + |y_i| \right) + \sum_{i \in \Omega_k^w} \xi_i^r (|x_i - x_k| + x_k) \right) \tag{3.1}$$

where $\xi_i^{c,a} = 1$ if traveler i has chosen mode **c** or **a** and 0 otherwise. Similarly, $\xi_i^r = 1$ if traveler i has chosen mode **r** and 0 otherwise.

One can define Λ_x and Λ_y so that $\Lambda = \Lambda_x + \Lambda_y$. Λ_y is independent of the design and travelers' distribution over the available modes. On the contrary, Λ_x depends on the design

and travelers' distributions over modes in the general case. However, when we assume that walking and vehicle attraction areas coincide, i.e., $\Omega_k^{\mathrm{c}} = \Omega_k^{\mathrm{w}}$, $\forall k \in [\![1, M]\!]$, Λ_x can be expressed without knowing the distribution of travelers over modes. Vehicle and walking attraction areas coincide under the condition $v_{\mathrm{st}}/u = v_{\mathrm{w}}/v_{\mathrm{r}} + 2T_{\mathrm{d}}v_{\mathrm{w}}/s_k, \forall k \in [\![0, M-1]\!]$.

$$\Lambda_x = \sum_{k=1}^{M} \left(\sum_{i \in \Omega_k^{\mathrm{c}+}} (x_i) + \sum_{i \in \Omega_k^{\mathrm{c}-}} (2x_k - x_i) \right) \tag{3.2}$$

where $\Omega_k^{\mathrm{c}+}$ is the part of Ω_k^{c} located upstream x_k and $\Omega_k^{\mathrm{c}-}$ is the part of Ω_k^{c} located downstream x_k.

From $M = 0$ to $M = 1$, Λ_x increases because one part of travelers, those within $\Omega_1^{\mathrm{c}-}$ go first upstream to join c_1, then travel downstream on the freeway or the MRT line. They experience a detour, bounded by $x_1 - X_{0,1}^{\mathrm{c}}$. When the spacing is constant, we have $s_k = B/(M+1)$, $k \in [\![1, M]\!]$. The sum of maximal detours for all attraction areas is noted D and defined by:

$$D = \sum_{k=1}^{M} x_k - X_{k-1,k}^{\mathrm{c}} \tag{3.3}$$

where $X_{k-1,k}^{\mathrm{c}}$ is well defined for $\frac{B}{M+1} > \frac{T_{\mathrm{d}} v_{\mathrm{w}} v_{\mathrm{r}}}{v_{\mathrm{w}} - v_{\mathrm{r}}}$. Replacing $X_{k-1,k}^{\mathrm{c}}$ by its expression (Equation 2.29) leads to:

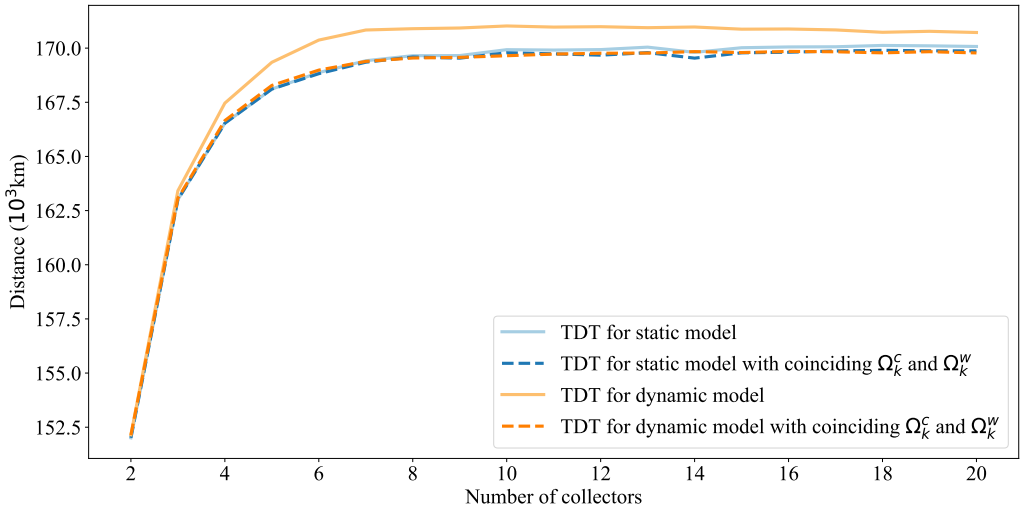
$$D = \frac{MB}{2(M+1)} \left(1 - \frac{v_{\mathrm{st}}}{u}\right) \tag{3.4}$$

Equation 3.4 indicates that the total area on which travelers undergo a detour ($D * W$) increases asymptotically. It justifies the profiles of Λ presented in Figure 3.2a for the monocentric city scenario. Static and dynamic models have almost overlapping curves with coinciding vehicle and walking attraction areas. Marginal differences are due to travelers aggregation in the static model.

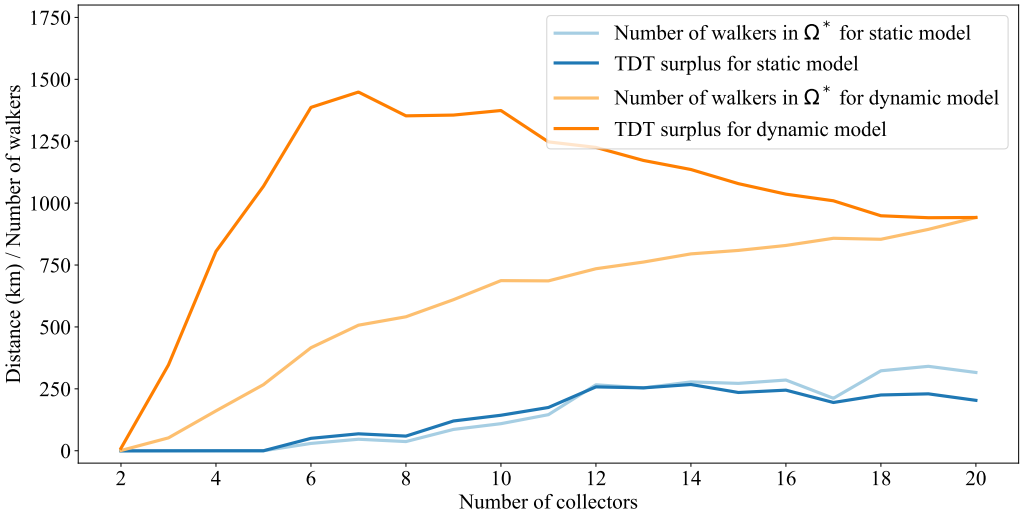
For non-coinciding vehicle and walking attraction areas, when $X_{k,k+1}^{\mathrm{w}} \leq X_{k,k+1}^{\mathrm{c}}$ (which is generally the case given the speed values), let us note:

$$\Omega^* = \bigcup_k \left(\Omega_k^{\mathrm{c}} \cap \Omega_{k+1}^{\mathrm{w}} \right) \tag{3.5}$$

A commuter $i \in \Omega^*$ choosing mode **r**, travels $2(x_{k+1} - x_i)$ more than if they have chosen mode **c** or **a**. The total additional distance traveled in the non-coinciding attraction areas configuration is closely related to the walk mode share within Ω^*. If their global profiles are analogous, the static and dynamic models slightly differ in evaluating this surplus, as shown in Figure 3.2b. The static model indicates fewer walkers in Ω^* and does not capture the long-distance walkers for $M \leq 5$, contrary to the dynamic model. The surplus does not deviate from the curve of the number of walkers in Ω^*. The dynamic model shows a concave surplus with respect to M. The best accessibility improvements, i.e., when the MRT attracts more travelers with, on average, a smaller walking distance, are made from $M = 2$ to $M = 7$.



(a) *Impact of the number of collectors on the TDT by all modes. The static and dynamic models lead to similar profiles with coinciding vehicle and walking attraction areas. For non-coinciding attraction areas, the static model misses the distance traveled by walkers departing from Ω^* compared to the dynamic model.*



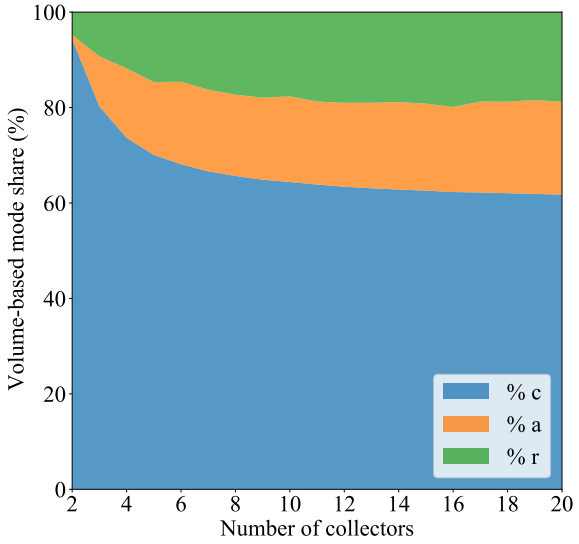
(b) *The static model shows fewer walkers departing from Ω^* and misses the impact of M on walking distances.*

FIGURE 3.2 – *Impact of the number of collectors with a constant spacing on the total distance traveled (TDT) for the dynamic and static models.*

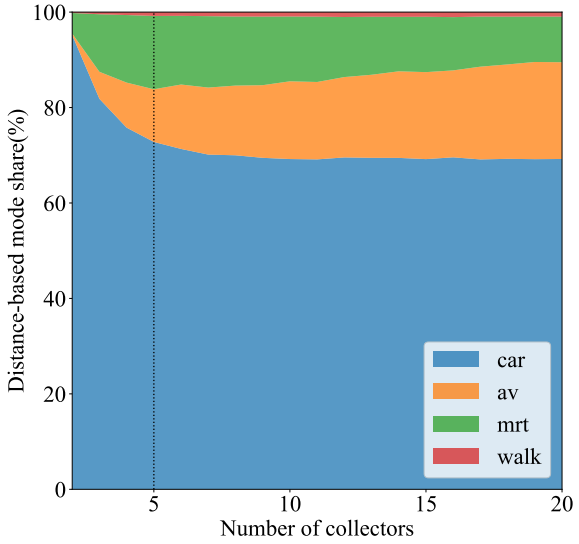
Both models are consistent regarding the total distance traveled. The marginal differences highlighted establishes the first asset of the dynamic model in richer effects of design on walking distances.

3.3.2 Cooperation and competition schemes

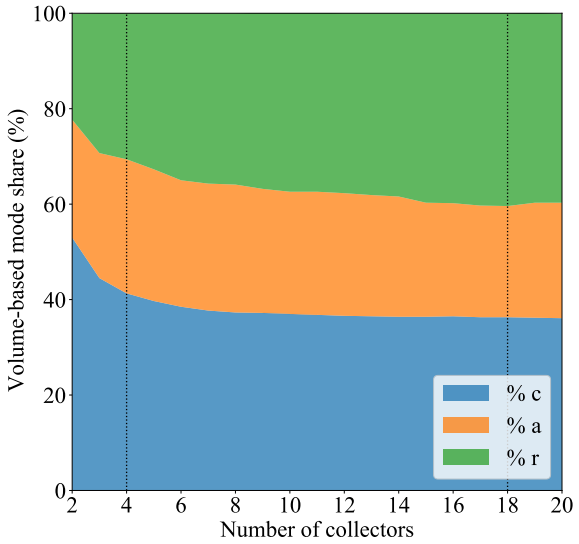
Figure 3.3 shows the volume-based and distance-based mode shares for the static and dynamic models for different M values.



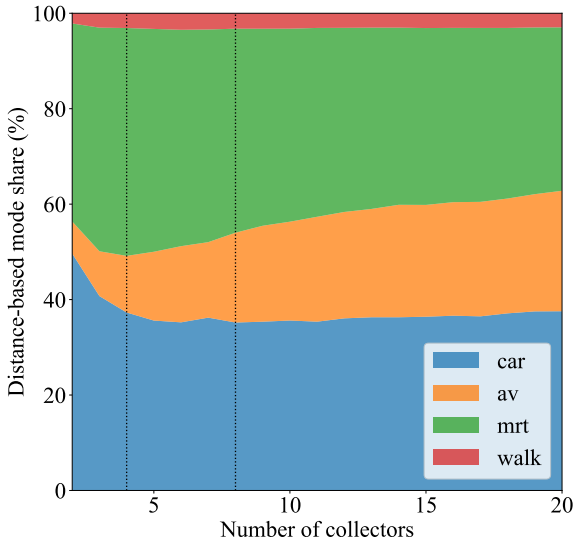
(a) *Volume-based mode share computed by the static model shows only one cooperation behavior between* ***r*** *and* ***a*** *versus* ***c****.*



(b) *Distance-based mode shares computed by the static model show cooperation then competition between* ***r*** *and* ***a****.*



(c) *Volume-based mode shares computed by the dynamic model show cooperation then competition between* ***r*** *and* ***a****.*



(d) *Distance-based mode shares computed by the dynamic model show cooperation followed by competition between* ***r*** *and* ***a****.*

FIGURE 3.3 – *Impact of the number of collectors with a constant spacing on volume-based and distance-based mode shares for static and dynamic models. The dashed vertical lines show the boundaries between two competition-cooperation schemes.*

They both present a decreasing **c** volume-based mode share profile. The most significant modal shifts from car to other modes are achieved for a small number of collectors. Over eight collectors, improving accessibility after adding a new collector has a limited impact on the volume-based mode share of **c**. The travelers who choose **c** whatever the number of collectors are "captives" to the car mode. Their number is directly linked to the CBD off-ramp capacity (μ_0) as **c** remains the most efficient mode of the system in free-flow conditions. As expected, the static model underestimates congestion and overestimates the volume-based

mode share of mode **c** compared to the dynamic model.

The models show different MRT-AMoD interactions. In terms of volume-based mode shares, the dynamic model highlights three competition-cooperation schemes over M (Figure 3.3c).

- $M \leq 4$: a new collector leads to a modal shift from **c** to **r** and **a**. MRT and AVs cooperate against cars.
- $4 < M \leq 18$: **r** benefits from a higher number of collectors and takes the market share from **c** and **a**. MRT and AMoD compete to the advantage of the MRT.
- $18 < M$: MRT and AVs still compete but **r** now loses the market share to **a**.

Opportunities to exploit **r**-**a** cooperation come for small numbers of collectors since competition starts from $M = 4$. The static model shows only the first of these schemes: MRT and AVs cooperate against cars. This behavior is stable as M increases. The static model does not identify competition between **a** and **r** in terms of volume-based mode share.

Distance-based mode share measures the usage of a mode regardless of the trip in which it intervenes. In terms of distance-based mode share, the dynamic model reveals three schemes:

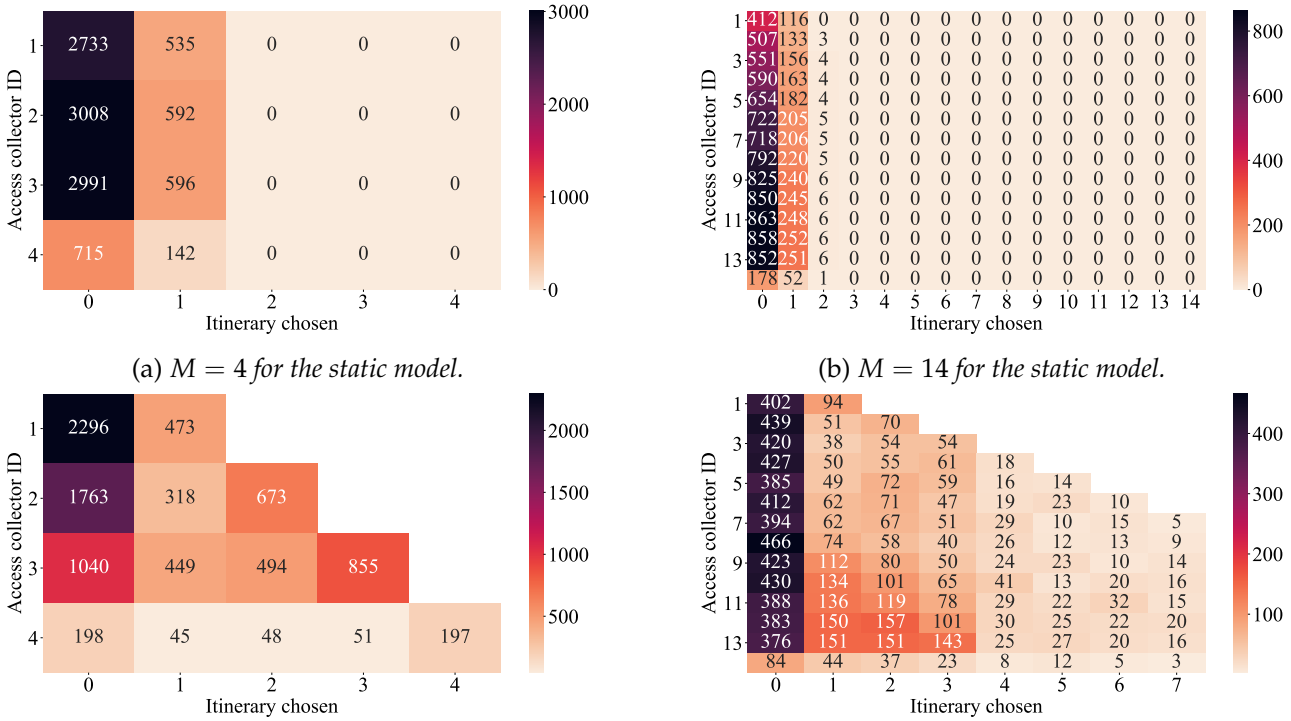
- $M \leq 4$: a new collector leads to lower car usage and higher MRT and AV use, and walking. Positive synergy between MRT, AV, and walking occurs in this scheme.
- $4 < M \leq 8$: **a** benefits from a higher number of collectors. AVs takes kilometers from **r**, and to a lesser extent from **c**. AVs compete mainly with the MRT and secondarily with cars.
- $8 < M$: MRT usage declines against both AVs and cars. The usage of AVs and cars grows significantly.

The static model captures only the first two schemes, as shown in Figure 3.3b.

Figure 3.4 confirms that the static model neglects the influence of design on travelers' distribution over itineraries. Travelers divert only via c_1, the closest collector to the CBD off-ramp, which is available for everyone. Collector c_2 is only used marginally for high values of M). For low M values, close and distant travelers load c_1 similarly. As M increases, the provenance of AV riders and car drivers progressively imbalances. Distant travelers load more c_0 and c_1 than close travelers. (Figure 3.4b). The car mode share is greater in upstream vehicle attraction areas, while the MRT mode share is greater in downstream walking attraction areas. Indeed, MRT suffers from an increasing cumulative time lost due to train dwelling when M grows.

The dynamic model leads to a more complex diversion pattern. Transfers take place at more than one collector. All the collectors from c_1 to c_7 are loaded when $M = 14$. The distribution of commuters over the transfer collectors is heterogeneous. As M increases, the distribution pattern evolves from the direct (Figure 3.4c) to the long first-mile (Figure 3.4d) type. In the direct first-mile pattern, most commuters in Ω_k^c request an AV to join c_k, their vehicle access collector. In the long first-mile pattern, more distant travelers use AVs to join

the most downstream collectors. The cooperation-competition schemes and distribution patterns are consistent. The direct first-mile pattern is a cooperation scheme, while the long first-mile pattern is a competition scheme.



(a) $M = 4$ *for the static model.*

(b) $M = 14$ *for the static model.*

(c) $M = 4$ *for the dynamic model.*

(d) $M = 14$ *for the dynamic model.*

FIGURE 3.4 – *Impact of the number of collectors with constant spacing on the distribution of commuters over itineraries for static and dynamic models. Only travelers that have chosen* ***c*** *or* ***a*** *appear here. The dynamic model shows an evolution in AMoD usage from a direct first-mile to a long first-mile pattern. The static model neglects this change of behavior.*

Considering dynamics makes it possible to capture the influence of design on the cooperation-competition schemes and distribution patterns.

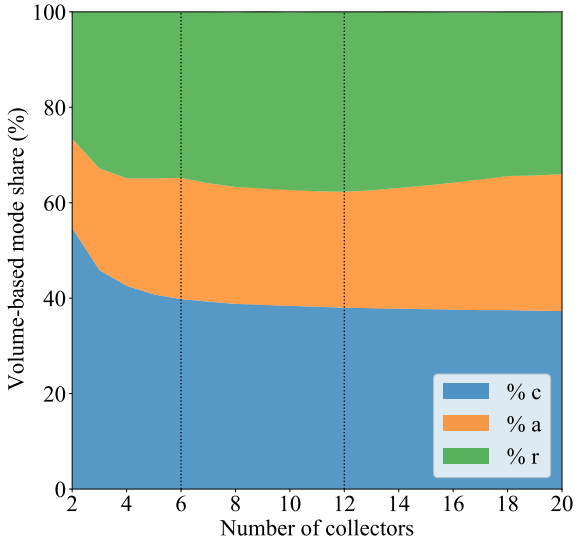
3.3.3 Analysis of the schemes in the light of DUE principles

To discuss the influence of design on the cooperation-competition schemes between the MRT and AMoD, we relaunch the sensitivity analysis of M with the pseudo-dynamic model, where T_s equals the mean service time obtained with the dynamic model. The schemes and patterns of Figure 3.5 are similar to those of Figures 3.3c, 3.3d, 3.4c and 3.4d.

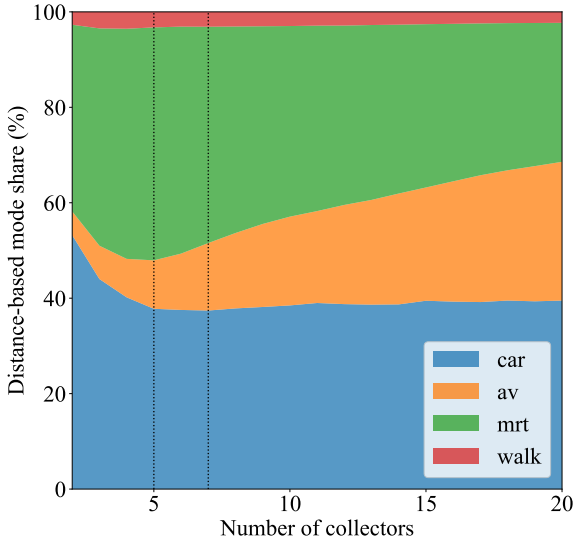
The boundaries between the schemes are different. Notably, the volume-based first competition scheme shortens. It extends from $M = 4$ to $M = 18$ with an endogenous time-variant service time and from $M = 6$ to $M = 12$ with a constant service time. A more significant distance-based modal shift occurs to the advantage of AVs. The lack of a feedback loop on service time favors AMoD.

Regarding the distribution of travelers over itineraries, the direct and long first-mile patterns are respectively more and less intense than in the dynamic model. For low M values,

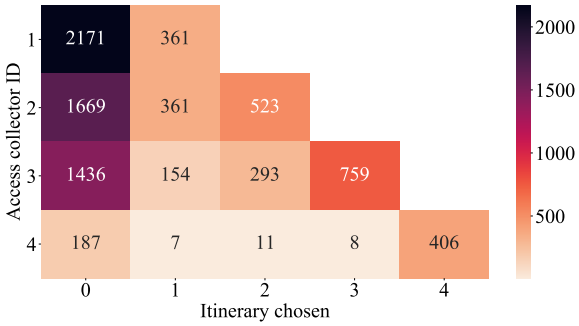
a larger share of AV riders transfer at their vehicle access collector. For high M values, the distribution of distant AV riders on downstream collectors is more spread.



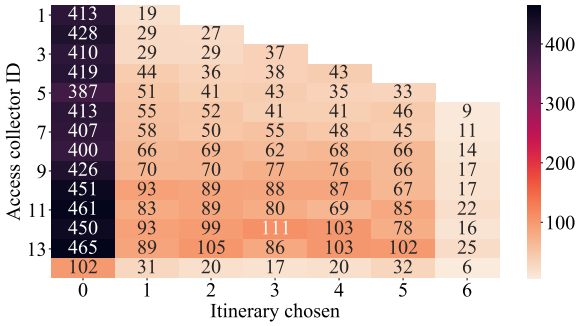
(a) *Volume-based mode shares with the pseudo-dynamic model show a smaller M range where* ***r*** *and* ***a*** *compete to the advantage of* ***r****.*



(b) *Distance-based mode shares with the pseudo-dynamic model show a greater modal shift toward AV.*



(c) *Itineraries for $M = 4$ show a more intense direct first-mile.*



(d) *Itineraries for $M = 14$ show a less intense long first-mile.*

FIGURE 3.5 – *Impact of the number of collectors with a constant spacing on volume-based, distance-based mode shares, and traveler distribution over itineraries for the pseudo-dynamic model.*

The UE principles presented in the previous chapter allow to better understand the pseudo-dynamic model outputs.

At $M = 2$, Ω_1^c is wide enough to saturate both μ_0 and μ_1: c_0 and c_1 are synchronized while c_2 is out of sync. Travelers in Ω_2^c divert through their vehicle access collector only: the direct first-mile pattern is almost perfect, with just 22 of them transferring at c_1.

From $M = 3$ to $M = 5$, the first $M - 1$ bottlenecks are synchronized. A smaller vehicle attraction area surrounds c_M. This bottleneck is used but not overloaded: $w_M(t)$ is null. It is out of sync with downstream bottlenecks.

As M increases, attraction areas narrow, first diversions start earlier, and additional travel time values are closer to each other, i.e., successive bottlenecks switch on more quickly. Trav-

elers are better spread on the $M-1$ first bottlenecks. The direct first-mile pattern here is less intense than in out of sync conditions.

At $M=5$, c_5 out of sync still guarantees that the more distant travelers follow the direct first-mile pattern. However, the acceleration of successive bottleneck activations and the reduced size of vehicle attraction areas generate the long first-mile pattern.

From $M=5$, adding a collector does not necessarily attract more diversions as the system has reached the maximum potential for diversion. The long first-mile pattern amplifies because more collectors are located close to the destination. These collectors attract fewer surrounding travelers and more upstream travelers who depart near bottlenecks with excessive additional travel times. Increasing M is similar to moving attractive diversion alternatives away from travelers on average. The car distance-based mode share increases during this scheme because $\mathbf{r}$ suffers from an increasing cumulative dwell time. Travelers departing from upstream attraction areas are more affected by the increase in cumulative dwell time than travelers departing from downstream attraction areas. Consequently, more drivers come from more distant vehicle attraction areas.

A good design should consider these dynamics and emerging behaviors regarding Ω_k^c size, the number of travelers diverting to $\mathbf{r}$, the maximum potential for diversion, and how additional travel times chain. The out of sync phenomenon forces the strict application of the direct first-mile pattern but might indicate that the overall diversion capacity is insufficient. A moderate long first-mile pattern is needed in practice to reach better car mitigation and MRT usage indicators.

3.4 The joint MRT-AMoD design problem in the West Lyon scenario

In the light of the analyses presented in section 3.3, we use the dynamic model in this section to address the MRT-AMoD design problem in a corridor where the many-to-one mobility pattern is problematic because unbalanced competition occurs between a freeway and an MRT line. While cars provide a convenient door-to-door trip, the MRT suffers from accessibility issues. Consequently, car usage is substantial, and the freeway is congested daily.

The transportation authority considers accrediting a private AMoD operator. The improvement of the MRT service (by extending the line, building new stations, or updating the train service pattern) and the regulation of AMoD are studied jointly to reduce car-use and improve MRT attractiveness while satisfying a performance criterion.

In this section, we may have several fleets of AVs. Each fleet operates within a coverage zone defined as a union of vehicle attraction areas. The fleet can provide a service to every commuter departing from its coverage area and necessarily drops them off at a collector included in this zone. A vehicle attraction area is not necessarily supplied by AVs: if the TA decides to forbid AMoD on Ω_k^c, then no AV will operate within this zone. Instead of m, we define the AVs ratio to specify the number of AVs operating within a zone per traveler departing from this zone.

3.4.1 Scenarios

3.4.1.1 The West Lyon corridor

The west part of the city of Lyon, France (Figure 3.6), is a relevant territory to apply our framework.

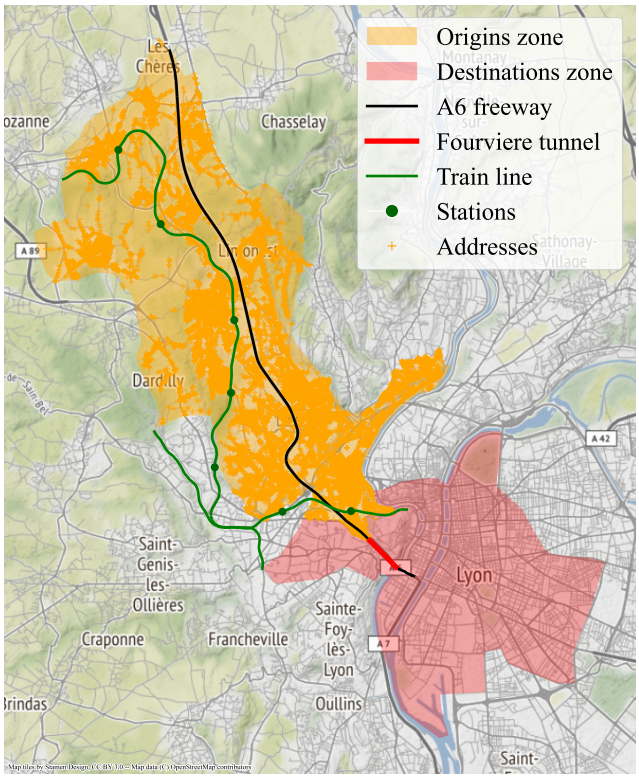


FIGURE 3.6 – *The Lyon corridor is composed of 9 towns surrounding the A6 freeway and the regional train line from Marcilly-d'Azergues to Lyon-Gorge-de-Loup.*

Freeway The A6 freeway links several suburban towns with Lyon city center. The Fourvière tunnel, the final part of the A6 freeway, is the entry point for the center. It acts as a bottleneck, causing significant congestion daily.

MRT A low-frequency regional train line operates along the corridor from Marcilly d' Azergues to Lyon Gorge de Loup. At Lyon Gorge de Loup station, travelers can access the Lyon meshed PT network (subway, buses). In practice, the train line is underused because of its low frequency (waiting time can be up to 30 minutes in the morning) and the lack of convenient access mode. Feeder buses operate in the downstream part of the corridor, but their routes are parallel to the train line. Their coverage is insufficient within this 5km-wide corridor. Moreover, its upstream part is outside the city's bus network. There are few regional buses in service.

Relevancy of the territory The context is favorable to AMoD deployment. Today, the share of the car mode using the corridor to the center is 64% versus 31% for PT (Urba Lyon, 2018).

At the local level, elected representatives request better first-mile solutions (Cerema, 2021). Regarding the metropolitan area, the policy goal is to improve the attractiveness of the train and change its role from a regional to an urban-suburban line. The public authorities have identified this corridor as an opportunity to develop new transportation systems. Several facilities favoring ridesharing have been deployed recently, including carpooling areas and an HOV lane. Entrepreneurs and research working on AVs have already launched pilot projects in Lyon Confluence (Navly) and Meyzieu (Mia). AMoD may contribute to rebalancing the mode shares in this corridor by providing an on-demand home-to-station service to commuters.

Realism of parameters Parameters for the West Lyon case study are available in table 3.3.

The demand scenario has been built on the Lyon Area OD matrix (2015) and BD TOPO addresses (2021). The first database allowed the computation of hourly trip generation rates for 5 hours in the corridor. We distributed departures according to a Poisson process in time and uniformly on addresses. In total, 14k commuting trips occur. Such a disaggregation process is sufficient to compare policies. A more precise design could be obtained using more detailed data sources that render the relationship between origin locations and departure times.

Our model requires the separation of cars and AV flows. This assumption is acceptable here as AVs could run on the HOV lane to escape the potential congestion spillback on the A6 freeway. Two lanes are available for all vehicles in the Fourviere tunnel. The demand considered represents only a part of the potential traffic passing through the tunnel. The capacity value chosen is half the maximal flow observed with inductive-loop detectors. The capacity value chosen for the other bottlenecks corresponds to 6 drop-off spots in front of a station and a drop-off time (T_f) of 30 seconds.

We assume a fixed headway that makes the MRT line attractive enough to justify the need for a feeder service. A 15 minute headway corresponds to the highest frequency currently operated.

Finally, to ensure the realism of the scenario, the number of AVs operating in the corridor is limited to 700, with an AVs ratio of 5%.

3.4.1.2 Investigated policies

We benchmark three distinct policies regarding AMoD deployment and three priority objectives for the TA. Protectionism consists in refusing to accredit any AV and focusing on MRT design only (no AMoD). Opportunism relies on AMoD regulation to foster cooperation between MRT and AVs (regulated AMoD). Liberalism focuses on MRT design while allowing the AMoD operator to serve its interest, i.e., seek increasing its profit (unregulated AMoD). Here, we consider a simple version of service pricing and cost schemes. The profit of AMoD is proportional to the amount of work achieved (in passengers * kilometers). Hence, AMoD operator's objective is reduced to maximizing the usage of AVs. The distance-based mode share associated with a mode measures its usage.

The priority objective of the TA may be to: maximize MRT usage, minimize car usage, or minimize the average travel time per traveler during the morning commute.

Table 3.4 presents the set of scenarios.

Scenario name	West Lyon
Description	Small suburban towns external to main urban area
Distribution of travelers	BD TOPO addresses
MRT type	Medium freq. regional trains
B * W **(km^2)**	20*5
Morning peak duration (h)	5 - 2015 Lyon Area OD matrix
Number of travelers	14k
v_{w} **(m/s)**	1.2
v_{st} **(m/s)**	14
u **(m/s)**	18
v_{r} **(m/s)**	25
h **(s)**	900
T_{d} **(s)**	45
μ_k**,** $0 < k \leq M$ **(veh/s)**	0.2
μ_0 **(veh/s)**	0.4
AVs ratio	5%
Nominal number of collectors	7
Nominal s_k**,** $0 \leq k < 7$ **(km)**	2,2,2,3,2,5,3
T_{s} **step size (s)**	150

TABLE 3.3 – *Parameters for the monocentric city scenario.*

3.4.2 Optimization framework

Figure 3.7 presents the complete framework and lists the possible design parameters.

Policy	Protectionism		Opportunism			Liberalism		
Regulation strategy	Forbid AVs		Liscense AVs by fleet and coverage zone			No regulation		
Design optimization	Unilateral		Joint			Consecutive		
Piority objective	Max MRT usage	Min avg. travel time	Max MRT usage	Min avg. travel time	Min car usage	Max MRT usage	Min avg. travel time	Min car usage

TABLE 3.4 – *Summary of the policy scenarios.*

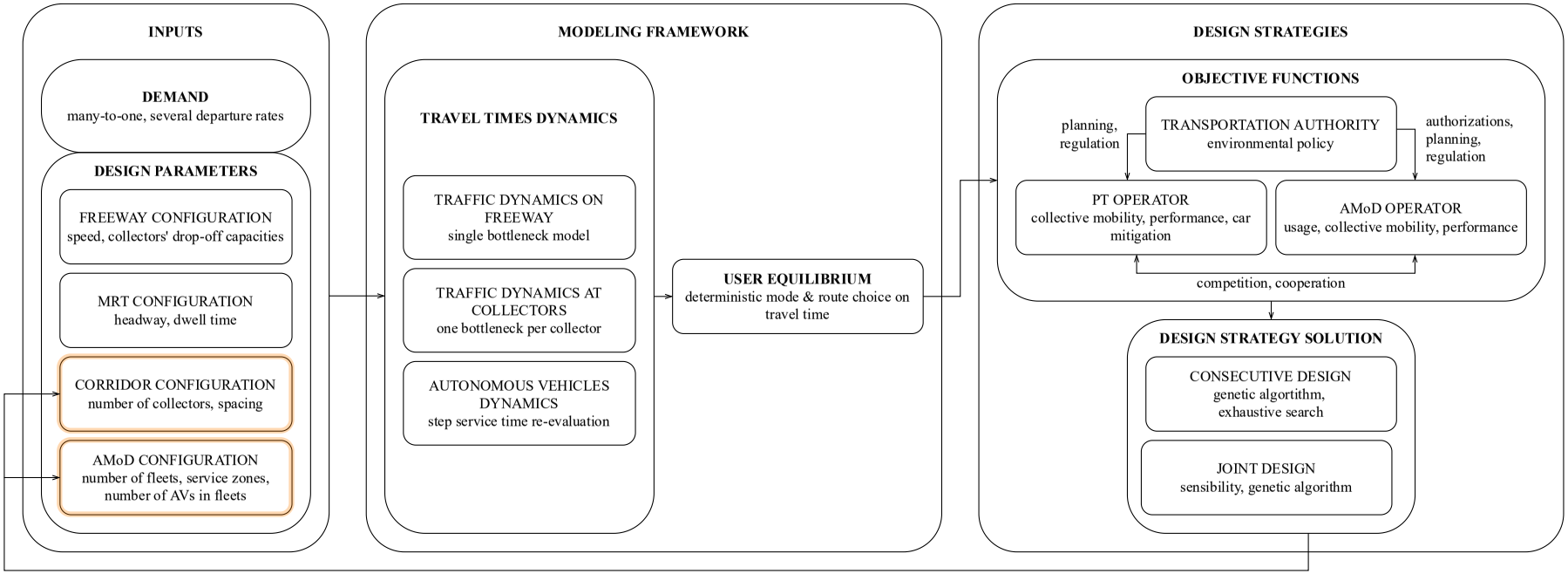


FIGURE 3.7 – *Optimization framework. Our dynamic modeling contribution linked with an optimization loop answers the MRT-AMoD design problem, with corridor and AMoD configurations as decision variables.*

3.4.2.1 Decision variables

This section justifies the choice for four of them as degrees of freedom: the number of collectors, their locations, the number of AV fleets, and their coverage zones.

Corridor configuration As shown in the monocentric city scenario, the number of collectors is a decisive parameter in fostering MRT-AMoD cooperation. If spacing was constant in section 3.3, it is not relevant for the West Lyon scenario, which has an inhomogeneous distribution of origins. Note that a lower bound on spacing should be imposed to prevent unrealistic configurations where two collectors are very close to each other.

AMoD configuration Imposing a boundary for one fleet coverage zone can mitigate the long first-mile pattern, as shown by the sensitivity analysis presented in Figure 3.8. Two groups of consecutive bottlenecks supplied by two different fleets stay out of sync, so the direct first-mile holds. Similar effects can be obtained by decreasing (and increasing respectively) downstream (and upstream respectively) the stations' drop-off capacities. However, this solution is less flexible than coverage zone geofencing, which could be demand-responsive and reviewed daily. Geofencing AMoD can have other positive effects such as service time reduction and reliability (Shen & Quadrifoglio, 2013), reduction of total distance traveled by empty AVs (Fagnant *et al.*, 2016), equity gains with more uniform service times over space (Gurumurthy *et al.*, 2021).

Fixed design parameters Freeway speed, bottleneck capacities, MRT headway, and AV ratio are other levers for lowering car performance but remain untouched in this study.

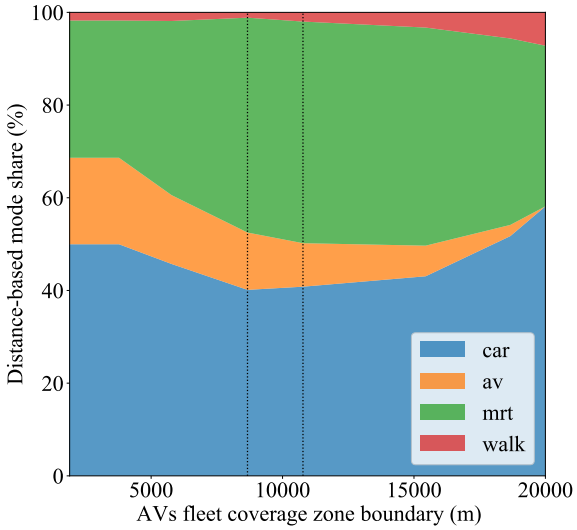


FIGURE 3.8 – *Sensitivity analysis of the downstream boundary of one fleet coverage zone on distance-based mode shares for the West Lyon scenario. Compared to a situation without AV, a single fleet reduces car usage, whatever its coverage zone, and improves MRT usage when its coverage zone includes up to 6 upstream collectors. Three schemes emerge: (i) MRT-AV cooperation versus cars for coverage zones extending from c_7 to at most c_4, (ii) AV competition with MRT and cars for coverage zone extending from c_7 to c_3, (iii) AV-cars competition with MRT for wider coverage zones.*

3.4.2.2 Optimization loop

Under protectionism, the TA optimizes MRT design unilaterally. Under liberalism, this optimization assumes that AMoD is composed of one fleet serving the whole corridor. Then, the unregulated AMoD adjusts its configuration to maximize its profit. Under opportunism, the TA jointly optimizes MRT and AMoD designs.

As the nature of the function that links our degrees of freedom with each objective is unknown (not necessarily convex nor linear), we choose to use a genetic algorithm (GA) when exhaustive research is not possible. Such a metaheuristic does not guarantee global optimality. Stopping criteria, such as the stability of the fitness function over a large number of iterations and the small number of distinct individuals in the current population, are signs that the algorithm has reached a local optimum. Local optimality is sufficient for the goal of this study to highlight the benefits of a joint MRT-AMoD design. Finding a good design instead of the optimal one still allows for assessing the lower bound of the gains achieved under each policy. Since a fleet coverage zone is compact and extends over one or several vehicle attraction areas, an exhaustive search for the optimal AMoD configuration is possible for $M < 10$.

Applying the GA to the MRT design requires discretizing the corridor into several location spots. A spot length is sufficiently large to have meaningful configurations (two freeway ramps/MRT platforms cannot be too close to each other) and sufficiently small to keep precise locations. Lower and upper bounds on the number of stations constrain the problem to account for the current corridor configuration and the finite investment budget. Then, the MRT design can be binary-encoded, where the chromosome length equals the number of location spots. Applying the GA to the AMoD design requires a ternary encoding where the chromosome length equals the number of collectors. For the joint MRT-AMoD design, the problem is quaternary-encoded, and the chromosome length equals the number of location spots. Figure 3.9 provides an example of encoding.

In terms of implementation, traditional genetic algorithm components are used. The roulette strategy with 1-elitism [1] selects individuals. Random pairs are formed within the set of selected individuals to become parents. The well-known two points crossover operation is executed with a probability of 0.5. To produce viable individuals, i.e., having a meaningful encoding and respecting the lower and upper bounds on the number of collectors, we use the efficient search for an entirely feasible crossover operator presented in Reid (1996).

Three types of mutation operators are defined for the corridor configuration. One consists in adding a new collector at a random free location spot. Another consists in removing a random collector. The last one shifts a random collector by Δx, where Δx is a random variable that follows a truncated normal distribution centered in the collector's current location. Accessible spots for the shifted collector are restricted to free spaces surrounding its initial location. One mutation operator is defined for the AMoD configuration. It consists in randomly changing the value of a gene at which a collector exists to produce a viable individual.

One mutation operator is applied to newborns with a certain probability at the begin-

[1] The best individual, i.e., the one with the highest objective function value, is directly copied into the next generation

ning. The mutation rate is adaptive to prevent an early homogenization of chromosomes. We increment the mutation rate when the standard deviation of the objective function values for all individuals in a generation is below a certain threshold. This evaluation is done every five generations.

Several population sizes have been tested. To start with and keep a diversified population in addition to limiting computing time, 80 individuals is a satisfying parameter.

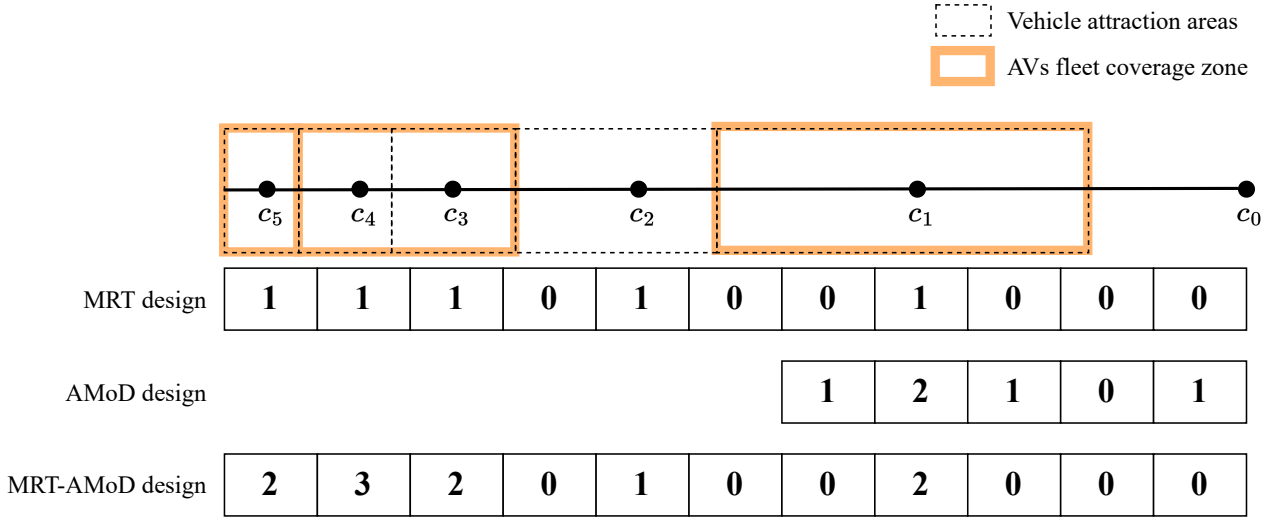


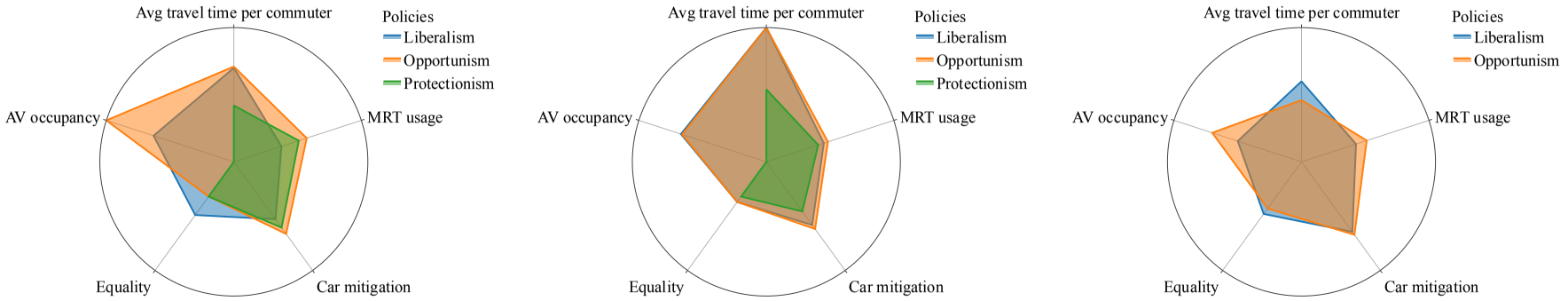
FIGURE 3.9 – *The MRT design is binary-encoded: 0 and 1 stand for an empty spot and a collector respectively. The AMoD design is ternary-encoded: 0, 1, and 2 stand for an unsupplied attraction area, the downstream extremity of a new coverage zone, and the continuity of a coverage zone respectively. The MRT-AMoD design is ternary-encoded: 0, 1, 2, and 3 stand for an empty spot, an unsupplied attraction area, the downstream extremity of a new coverage zone, and the continuity of a coverage zone respectively.*

3.4.3 Results

Figure 3.11 presents the best designs found by the GA. Table 3.5 gathers the numerical results, and Figure 3.10 compares the scores obtained by each policy regarding five indicators: the average travel time per commuter (including those using streets only and located between c_1 and destination), the MRT distance-based mode share, the car mitigation (sum of MRT, AVs and walking distance-based mode shares), the average number of commuters served per AV, and the Gini coefficient of travelers' waiting times (including service time and waiting time at bottlenecks).

Scenario	**West Lyon**							
Priority objective	Maximize MRT usage			Minimize average travel time			Minimize car usage	
Policy	Protectionism	Opportunism	Liberalism	Protectionism	Opportunism	Liberalism	Opportunism	Liberalism
M	15	9	6	17	13	12	5	10
Number of fleets	0	8	1	0	2	2	1	1
c volume-based mode share	51.1	48.7	48.6	51.2	49.0	49.6	48.9	48.5
r volume-based mode share	48.9	5.0	9.3	48.8	0.0	0.0	25.2	12.9
a volume-based mode share	0.0	46.3	42.1	0.0	51.0	50.4	25.9	38.5
Car distance-based mode share	39.4	33.6	47.1	54.3	38.1	41.8	32.8	35.5
MRT distance-based mode share	51.0	57.2	37.4	40.7	48.3	45.1	51.2	43.8
AV distance-based mode share	0.0	8.2	14.6	0.0	13.7	13.1	8.6	19.1
Walking distance-based mode share	9.5	1.0	0.9	4.9	0.0	0.0	7.4	1.6
Avg. TT per trav.	36:37	21:45	22:13	28:23	15:28	15:31	33:21	25:46
Gini coef.	0.68	0.68	0.51	0.68	0.63	0.63	0.57	0.52
AV occupancy	-	15.6	9.8	-	10.3	10.4	10.9	7.8

TABLE 3.5 – *Results for West Lyon.*



(a) *Maximize MRT usage objective.* (b) *Minimize average travel time objective.* (c) *Minimize car usage objective.*

FIGURE 3.10 – *Scores depending on priority objective and policy for the West Lyon corridor.*

3.4.3.1 Maximize MRT usage

When the priority of the TA is to maximize MRT usage, opportunism performs better than other policies regarding all the indicators except delays uniformity (Figure 3.10a). The MRT distance-based mode share reaches 57.5%, which is 6% more than under protectionism, and 20% more than under liberalism. The sum of AV and walking distance-based mode shares is similar in the protectionism and opportunism scenarios (9.5%). It increases by 6% under liberalism, indicating the presence of long first-mile rides. These long AV trips compete with the MRT rather than the car: the car mitigation score of liberalism is lower than protectionism. Liberalism leads to counterproductive designs.

However, the MRT design found under liberalism is the closest to the current one (Figure 3.11d). It follows the urban geography, with one station per town. The deployment of AVs and the refinement of the MRT service pattern (skip stops c_2 and c_3) can bring benefits to commuters, notably lower travel times on average and more uniformly distributed waiting times.

The liberalism scenario with the objective of maximizing MRT usage reaches the highest score regarding delay uniformity. Few travelers experience a null delay. Commuters departing from the downstream part of the corridor are subject to waiting times at the CBD off-ramp, like those departing from the upstream part which AMoD does not supply. Commuters departing from the middle experience non-null AMoD service time and AV-to-MRT transfer time.

On the contrary, 20% of travelers undergo 70% of the total waiting time under opportunism. Drivers mainly come from the downstream part of the corridor and undergo considerable w_0. As AMoD splits onto small coverage zones, T_s and w_k are null or small for those diverting to **a**.

Protectionism reaches the same score as opportunism regarding delay uniformity. MRT riders are not subject to any delay, contrary to drivers.

Under opportunism and liberalism, the optimization of design led to substantial s_2 (spacing between c_1 and c_2). This inter-station plays a considerable role in the distribution of travelers over itineraries since expanding Ω_1^c results in overloading bottleneck μ_0 with travelers close to the CBD. In this way, long-distance trips are subject to diversion.

A one fleet per collector strategy emerges from the joint optimization of MRT and AMoD designs, as shown in Figure 3.11a. This strategy avoids the long first-mile pattern since bottlenecks remain out of sync.

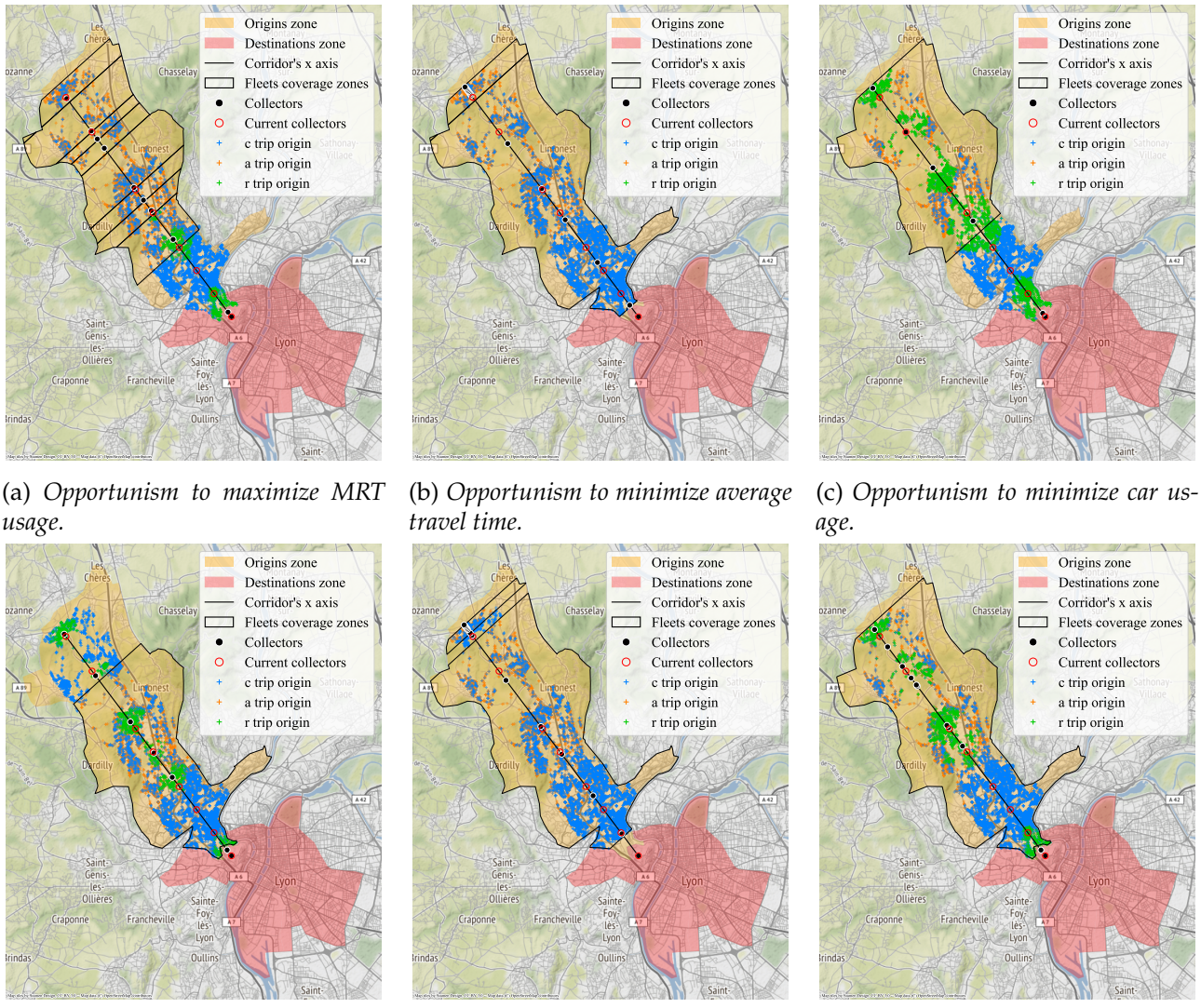
Moreover, it allows keeping short service times. The service times of all the fleets except those serving Ω_2^c and Ω_8^c remain shorter than 150s. In Ω_2^c, travelers close to c_2 prefer to walk while travelers far from c_2 request an AV. Since relocation and serving efforts are substantial, the service time in Ω_2^c is longer than 150s. Similarly, relocation and serving distances are long in Ω_8^c because travelers' origins are far from c_8. The attraction area extends over a sparsely populated territory between the towns of Lissieu-Dommartin and Civrieux d'Azergues.

Finally, it improves the AV utilization rate. An AV serves 5 customers per hour on average under opportunism against 3 customers per hour under liberalism. The indicator is homogeneous across fleets, except for the fleet surrounding c_8, which has a slightly smaller

score (12.8 customers).

The geometry of coverage zones under liberalism is different. One fleet covers a huge downstream part of the corridor. The AMoD operator deploys one fleet over a wide area to increase AVs usage. In this way, it takes advantage of the synchronization between bottlenecks and can exploit the total capacity of each **a** diversion itinerary. The exclusion of attraction areas where the market share of AMoD is limited (here Ω_5^c and Ω_6^c) allows it to keep its service time as small as possible in other areas.

Although protectionism performs well for maximizing MRT usage, it has the worst average travel time of all the scenarios (36:37). AMoD deployment leads to a 15min reduction, whether AVs are regulated or not.



(a) *Opportunism to maximize MRT usage.*

(b) *Opportunism to minimize average travel time.*

(c) *Opportunism to minimize car usage.*

(d) *Liberalism to maximize MRT usage.*

(e) *Liberalism to minimize average travel time.*

(f) *Liberalism to minimize car usage.*

FIGURE 3.11 – *Resulting designs for opportunism and liberalism scenarios.*

3.4.3.2 Minimize average travel time

When the main objective is to minimize the average travel time, liberalism and opportunism obtain similar results and outperform protectionism for all the indicators (Figure 3.10b). Deploying AVs reduces the average travel time by 46%, reaching the value of 15min30s. Regulation is not necessary here: improving performance comes down to increasing AMoD usage. MRT has been designed considering that one fleet serves the whole corridor, which is close to what emerges through the profit maximization process by AMoD (Figure 3.11e). The distance-based mode share of the car (and MRT respectively) is slightly smaller (and higher respectively) under opportunism. Although liberalism provides satisfying results, a joint design helps to reach an even better equilibrium that benefits travelers, AMoD, and the TA.

The positioning of collectors is similar in both scenarios, except for c_1, which is closer to the destination under opportunism (Figure 3.11b). Thus, more downstream travelers are included in Ω_1^{c} and contribute to congestion at μ_0. This explains that opportunism performs slightly better in car and MRT distance-based mode shares than liberalism. The AMoD configurations obtained both have one downstream fleet whose coverage zone ends just before Civrieux d'Azergues (the most upstream town).

Since $v_{\mathrm{r}} > u$ here, $\Delta_{i,p}^{\mathrm{a}}$ decreases as p increases. All additional travel times are positive given the substantial headway, but it would be better for traveler $i \in \Omega_k^{\mathrm{c}}$ to divert in priority as far upstream as possible, i.e., at c_k. The order of additional travel times favors the direct first-mile pattern. The long first-mile pattern has almost entirely disappeared. Travelers use AV to join one station downstream of their access collector at most.

Consequently, service time remains below the discontinuity threshold, and additional travel times by AVs are independent of i ($\Delta_{i,p}^{\mathrm{a}} = \Delta_p^{\mathrm{a}}$). The collectors are positioned so that all Δ_p^{a} are very close to each other. Then, **a**-div1 starts on all the collectors supplied by AVs nearly simultaneously. This occurs before **r**-div1, so the share of the walking distance-based mode is null. Bottlenecks are synchronized until the unloading phase. The design allows the system to enter as soon as possible and remain in a global **a**-div1 state. As a result, w_0 and $w_k (k \in [\![1, 5]\!])$ remain small and stable (5min for w_0 and 60s for w_k).

Upstream, around Civrieux d'Azergues, all the designs share an accumulatioapan of collectors. For scenarios with AVs, among these collectors, none is used as an access point by walking, and only the most downstream one is chosen for transfer from AV to MRT. Consequently, no traveler suffers from the significant cumulative dwell time due to this accumulation of collectors. It is a trick found by the GA to improve the travel times of drivers and AV riders in accessing the freeway. Keeping only the most downstream of these collectors leads to a marginal increase in the average travel time per commuter.

3.4.3.3 Minimize car usage

When car mitigation is the main objective, opportunism once again obtains the best score among all the scenarios (Figure 3.10c). It relies more on walking than liberalism (+5.8%). This explains the higher average travel time score (+8min).

Liberalism performs worse than the "opportunism to maximize MRT usage" scenario in

terms of car mitigation. However, the resulting AMoD configuration is precisely the one considered by the TA to design the MRT line (Figure 3.11f). It is a sign that MRT-AMoD cooperation is necessary to take the market share from **c**, especially concerning the drivers that cannot be attracted to **r** whatever the MRT design.

3.5 Conclusion and discussion

In this chapter, the comparison of the models introduced in chapter 2 showed that considering dynamics allows capturing richer cooperation-competition schemes between MRT and AMoD. Two extreme patterns emerged from individual choices depending on the MRT design. In the long first-mile pattern, travelers favor downstream stations to transfer, so the AV legs of trips are long. In the direct first-mile pattern, travelers tend to use AVs for smaller distances to join their access collector and transfer there. These patterns were explained in the light of UE principles.

The application of our model to the West Lyon corridor showed that the joint design of MRT and AMoD can foster their cooperation. Indeed, opportunism provided the best results for each priority objective of the TA. A single fleet per collector design was found in the West Lyon corridor when the objective was to maximize MRT usage. However, such regulation offered travelers fewer choices for diversion, less flexibility, and uniformity in the delays experienced. Liberalism was less reliable in achieving the given objectives but ensured greater uniformity of delays. Protectionism performed adequately depending on the distribution of origins but never obtained the highest scores.

The one fleet per collector design strategy emerging from the opportunism policy is consistent with the paradigm of Autonomous Mobility District (AMD). AMD refers to using AV technology in a limited geographical area that generally includes an MRT station. AMD could overcome several urban planning issues. Hou *et al.* (2018) listed the assumed benefits of an AMD, including reducing car use, parking lots, and pedestrian-oriented land use. Many publications have studied intra-AMD mobility with simulation (Huang *et al.*, 2021, Shen *et al.*, 2018, Scheltes & de Almeida Correia, 2017), but few have studied inter-AMD mobility.

A strong limitation of our application is the reality gap that stems from the deterministic route choice based on travel time only. We did not include the monetary aspect because it involves making additional assumptions regarding AMoD and MRT fare schemes. Such assumptions prevent us from accessing the primary cooperation-competition schemes between MRT and AMoD. We did not include a transfer penalty to account for the discomfort of changing mode because the transfer between AV and MRT is already penalized by a wait time for drop-offs. Adding a constant transfer penalty to $T^{a}_{i,p}$ takes the form of additional travel times $\Delta^{a}_{i,p}$ and changes the times at which the system changes state but does not modify the states themselves. In order to address the reality gap in the West Lyon corridor, future research could evaluate the designs found in this study through simulations on agent and activity based platforms, which are more accurate for reproducing travelers' choices and AMoD operation.

The use of GA to solve the optimization problem is quite direct, does not require any

knowledge about the property of the objective function (GAs are good for exploring a search space with no a priori), and is highly flexible. Refining the dynamic model is possible without changing the implementation of the heuristic. Adding new decision variables is possible as long as one finds a proper encoding and the problem size remains compatible with the simulation budget. Indeed, a considerable number of simulations should feed GAs. Moreover, some types of constraints applying to the problem can be directly integrated into the encoding, mutation, and crossover operators, as in our case. In contrast, other types of constraint require additional simulation time to be checked and lower the algorithm's performance. GAs are sensitive to meta-parameters such as the population size, the mutation rate, etc. Above all, they do not guarantee the optimality of the solution found. Here, we obtained sufficiently satisfying results and did not try other optimization methods. However, more experiments would be required to ensure the robustness of the framework for future exploitation.

Finally, we have restricted the study to optimizing four types of decision variables. Other parameters of our dynamic model could become relevant regulation levers, such as bottleneck capacities and AVs ratio. Moreover, one can question the validity of the conclusions presented in the case of West Lyon for other types of corridors. We did not include the additional tests launched on different theoretical corridor types to keep this chapter focused and readable. The reader can find them in Appendix A.1.

Conclusion of Part II

In this part, we addressed the joint MRT design and AMoD regulation for the morning commute in a corridor. The design of a system is usually discussed within the most demanding conditions (during rush hours). We have extensively simplified these conditions here. In practice, the demand pattern in a corridor is more complex. The models presented focus on the main flow of travelers during the morning commute for a monocentric city. The design is realized by considering this main flow only and can be disadvantageous for secondary flows. The method proposed is fast to highlight good practices in terms of design depending on the regulator objective (e.g., one fleet per collector, no supply on the most downstream part of the corridor, etc.). However, it implies some preliminary and extra steps in practice.

The regulator needs first to identify the relevant corridors surrounding the city. The question of the study case boundaries is out of the scope of this thesis, but identifying the catchment area of the freeway and MRT line may be tricky. Good field knowledge and using travel survey data and job mobility data are recommended for this step. Also, decision-makers should check the compatibility of the territory reality with the implications of some assumptions we formulated. Notably, the non-interaction of cars and AVs flows implies that (i) congestion downstream the freeway does not interfere with cars and AVs entry and exit from the freeway and that (ii) cars can use different roads than AVs to access the freeway near MRT stations.

Before its implementation, the design should be evaluated more precisely with the entire demand and the actual network. Our approach can work in combination with microscopic simulation for example.

Scaling up the methodology to a broader level would require additional work. For polycentric or twin cities, we could use the same framework within a corridor linking two centers by considering their flow exchange. The analytical solution to the DUE under two crossed one-to-one patterns should be similar to the one presented in Chapter 2 until two neighboring bottlenecks are activated and congested by different flows: one bottleneck by travelers from the left-to-right flow, the other bottleneck by travelers from the right-to-left flow. The conditions under which the left and right groups of bottlenecks synchronize and the next steps of the DUE should be specified. Additional work is needed to apply the same modeling steps in this case, namely, disaggregating travelers and computing AMoD service time endogenously. However, we believe the tools developed in the many-to-one corridor case can scale to a "many-to-two" corridor.

Yet, they cannot deal with more complex demand patterns where primary and secondary traveler flows interact. Indeed, the stakes change then, and some modeling choices formulated in this part become irrelevant (e.g., point congestion instead of spread congestion). Consequently, we move to another approach in the next part to treat more realistic demand patterns and include travelers' specificities more easily.

Part III

Designing a pricing scheme in a large urban area case study

4. AMoD operational policy modeling and analysis

The elements of this chapter have been presented in:

- Cortina, M., Chiabaut, N., Leclercq, L. (2023). An assignment-based dispatching heuristic for agglomeration-scale AMoD: sensitivity analysis of the demand prediction horizon on the benefits for supplier and customers, In *Transportation Research Board 102nd Annual Meeting*, Washington DC, USA.
- Cortina, M., Chiabaut, N., Leclercq, L. (2023). Equity of a batch-matching on horizon policy for Autonomous Mobility on Demand, In *24ème Congrès Annuel de la Société Française de Recherche Opérationnelle et d'aide à la décision*, Rennes, France.

4.1 Motivation and objectives

In part III of this thesis, we seek to design a pricing scheme for encouraging cooperation between PT and AMoD. As explained in chapter 1, we choose to address the problem through agent-based simulation in order to capture the individual behaviors of travelers and AVs. This approach requires defining the strategies commuters and AMoD operator adopt to pursue their respective objectives. In the system, travelers are selfish agents who aim to minimize their travel costs. The AMoD operator manages a fleet of AVs in a centralized manner over a large urban area to maximize its profit.

In this thesis, we assume that the latter agent has no long-term economic strategy (e.g., offering very cheap rides at a loss to catch a more significant part of the mode share and then increasing fares to make more profit). Consequently, its operational policy is its primary lever to maximize its profit. When the number of AVs in the fleet only allows answering a part of requests [1], the operational policy guides AVs toward the most profitable rides. The level of service in the different zones of the urban area derives from the operational policy,

[1] In part III of this thesis, for the design of monetary regulations, we assume a fixed fleet size. Depending on the values of regulation levers, the fleet can be undersized, oversized, or correctly sized. In this chapter, under no regulation, the fleet is slightly undersized and spatial disparities in the level of service provided by AMoD appears. Existing TNCs, such as Uber and Lyft, tend to have different policies: they attract as many drivers as possible to have an oversized fleet able to meet the whole demand with acceptable waiting times. In the context of AMoD we consider, the stakes are different since the AMoD operator initially invest money to buy AVs. For this reason, the fixed undersized fleet size assumption is reasonable.

so as travelers' choices between driving their car, riding an AV for a door-to-door trip, or only for the first-mile leg of their trip. There is a link between AMoD operational policy and cooperation between PT and AMoD. The sensitivity of the pricing scheme on the system depends on this policy. AMoD operational policy is a crucial system element that deserves to be described and analyzed.

The first objective of this chapter is to describe and justify the model chosen for AMoD operational policy. We present four variants of a batch-matching over horizon policy.

In the context of the pricing problem, the demand knowledge horizon is necessary. An instantaneous matching strategy (without rebalancing) does not enable the AMoD operator to anticipate far requests that benefit from a subsidy, for example. Without a horizon, the operator cannot optimize its fleet according to the taxes and subsidies the regulator proposes. Hence, we suppose that demand is known over the horizon. Batch-matching usually performs with a batch of requests issued in a short period (a few minutes at most). The second objective of this chapter is to highlight the range of horizon length on which the batch-matching approach is relevant. We describe the fleet behavior depending on the policy variant and the horizon length in a theoretical scenario resembling the scenario we tackle in the next chapter. This study allows us to find, for this scenario, the policy variant and the horizon length allowing the operator to manage the fleet in the best way, i.e., make a maximum profit while staying tuned to eventual taxes and subsidies waived by regulator.

In the same scenario, for the selected variant and horizon length, we check the sensitivity of a naive pricing scheme on the system behavior, especially on travelers-side indicators. We check that the system expectedly reacts to the prices. This evaluation loops the evaluation of the operational policy in the theoretical urban area.

Finally, the last objective of this chapter is to discuss the limitations of the proposed approach for the more realistic scenario tackled in chapter 5, with a more significant number of requests and AVs.

4.2 Notations for this chapter

TABLE 4.1 – *Notations for chapter 4.*

Notation	Definition
t	Time
H	Horizon length
R	Set of requests AMoD operator knows at t
R_{O}	Set of open requests AMoD operator knows at t
e_r	Earliest pick-up time for request $r \in R$
l_r	Latest pick-up time for request $r \in R$
w_r	Maximum duration the traveler associated with request $r \in R$ can wait to be picked up
V	Set of AVs v composing AMoD fleet

Continue on the next page

TABLE 4.1 – *Notations for chapter 4 (continued).*

Notation	Definition
K	Maximum number of requests an AV can have in its service plan to be considered available
R_v	Set of requests that have been permanently matched with AV v
V_{O}	Set of available AVs at t
P_{R_v}	Service plan of v, it is the ordered set $\{r_1, ..., r_k\}$, with $k \leq K+1$
r_i	Request located at index i of P_{R_v}
α	Parameter for the rolling horizon principle
p_r	Gross income an AV can earn by serving request r
f_0^{a}	Base fare of AMoD fare scheme
$f_{\mathrm{d}}^{\mathrm{a}}$	Distance-based fare of AMoD fare scheme
$f_{\mathrm{t}}^{\mathrm{a}}$	Time-based fare of AMoD fare scheme
d_r	Distance on the shortest path between request r pick-up and drop-off points
t_r	Travel time on the shortest path between request r pick-up and drop-off points
$u_{v,r}$	Utility of the match between AV v and request r
$u(P_{R_v})$	Utility of the service plan of AV v
$P^*_{R_v \cup \{r\}}$	Best service plan for v including all requests in R_v and the new request r
$\mathrm{C}_{\mathrm{d}}^{\mathrm{a}}$	Distance-based operation cost of an AV
d_{r_{i-1},r_i}	Distance of the shortest path between r_{i-1} drop-off point and r_i pick-up point
d_{r_0,r_1}	Distance on the shortest path between AV current location and the pick-up point of the first request in its service plan
$E_{r_i\|P_{R_v}}$	Time at which v arrives at r_i pick-up point following its service plan
$L_{r_i\|P_{R_v}}$	Time at which v leaves r_i pick-up point following its service plan
$\xi_{r_i\|P_{R_v}}$	Binary variable that checks if time window of r_i is satisfied for service plan P_{R_v}
t_{r_0,r_1}	Travel time on the shortest path between AV current location and r_1 pick-up point
t_{r_1,r_0}	Travel time on the shortest path between r_1 pick-up point and AV current location
$t_{r_i,r_{i+1}}$	Travel time on the shortest path between r_i drop-off point and r_{i+1} pick-up point
$x_{v,r}$	Binary variable which equals 1 if v is matched with r, and 0 otherwise
$G = (N, A)$	Digraph representing the multimodal network with N the vertices and A the arcs
$G_m(N_m, A_m)$	Layer of the digraph G corresponding to mode $m \in \{\mathrm{w}, \mathrm{c}, \mathrm{a}, \mathrm{r}, \mathrm{s}, \mathrm{b}\}$ for walk, car, AV, train, subway and bus
v_m	Cruising speed of mode $m \in \{\mathrm{r}, \mathrm{s}, \mathrm{b}\}$
h_m	Headway of mode $m \in \{\mathrm{r}, \mathrm{s}, \mathrm{b}\}$
τ_m	Time lost per station for mode $m \in \{\mathrm{r}, \mathrm{s}, \mathrm{b}\}$

Continue on the next page

TABLE 4.1 – *Notations for chapter 4 (continued).*

Notation	Definition
τ_{a}	Reference AMoD pick-up time
$C_{\mathrm{d}}^{\mathrm{c}}$	Distance-based cost of car
K^{c}	Parking capacity
p^{c}	Parking price
β_a	Value of time for activity $a \in \{\mathtt{ride}, \mathtt{drive}, \mathtt{wait}\}$
$\gamma_{\mathrm{transfer}}$	Transfer penalty
v_{U_k}	Mean speed in urban zone U_k
V_r	Set of AVs candidate to be matched with r in the oversupply mode of the event-based matching heuristic
R'_v	Set of requests candidate to be matched with v in the undersupply mode of the event-based matching heuristic

4.3 AMoD operational policy

This section treats the first objective of this chapter, namely, describing the model chosen for AMoD operational policy.

4.3.1 Batch-matching over horizon

4.3.1.1 Matching, routing, rebalancing

The operational policy comprises up to four components as shown in Figure 4.1: order matching, routing, rebalancing, and ride-sharing (Zardini *et al.*, 2022). Matching refers to the assignment of customer requests to AVs. The fleet manager looks for the best matching regarding its objective. Once an AV has been assigned a list of requests to serve, routing builds an optimal route passing through pickup and dropoff points and maximizing AMoD operator objective while respecting travelers' constraints. Rebalancing corresponds to the relocation of empty AVs from oversupplied to undersupplied zones. It aims at minimizing the imbalance between supply and demand due to asymmetric mobility patterns. Finally, the three other tasks become more complex when ride-sharing is enabled. The ride-sharing component achieves matching and routing, assuming that several requests can be handled simultaneously by a single AV.

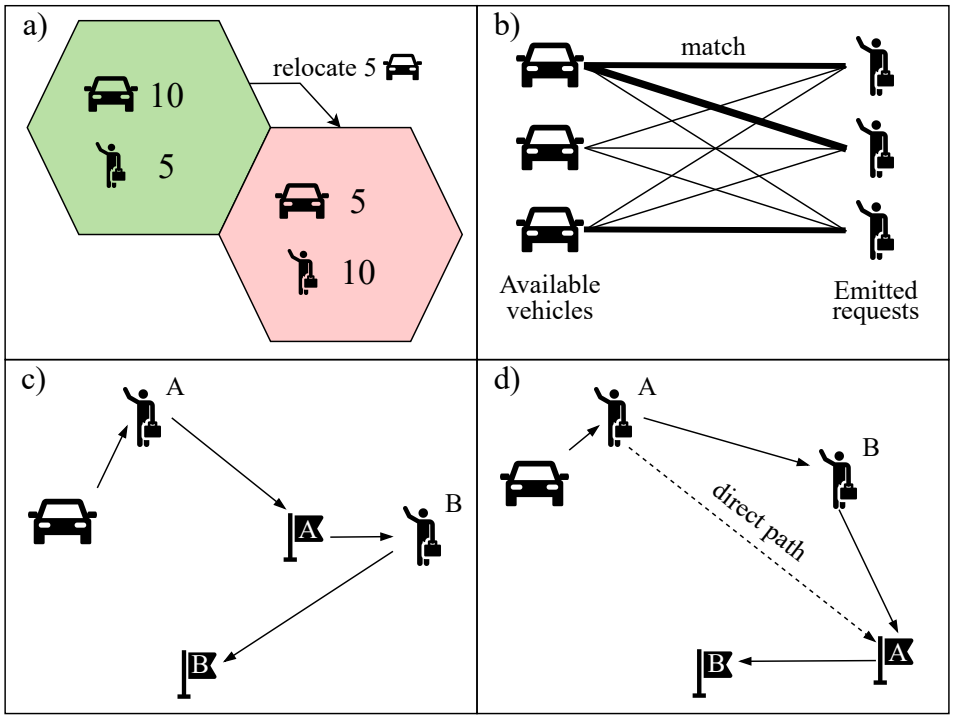


FIGURE 4.1 – *Components of AMoD operational policy. (a) Rebalancing, (b) Matching, (c) Routing, (d) Ride-sharing.*

4.3.1.2 Literature review

When it comes to modeling the operational policy for a fleet of vehicles, the first question is how dynamic the environment is. The answer conditions the type of formulation and

resolution approach that can be adopted.

In static environments, matching, routing, rebalancing, and, eventually, ride-sharing are usually achieved jointly. An external component can address the ride-sharing part before solving the matching, routing, and rebalancing problems, by grouping the compatible requests through clustering methods for example (Veve, 2021). An environment is said to be static when all the requests addressed to the service during the studied period (typically, a period of the day) are known in advance. It concerns reservation-based services for example (Ma *et al.*, 2017, Pimenta *et al.*, 2017). A state-of-the-art approach to deal with static environments is formulating the case as a one-to-one pick-up and delivery problem (PDP). In PDP, a set of routes should be built to satisfy several transportation requests while maximizing or minimizing a certain objective function (Savelsbergh & Sol, 1995). In one-to-one PDP, each request has a single pick-up point and a single drop-off point. When travelers send requests, we speak of Dial-A-Ride Problem (DARP). The exact solution to a multi-vehicle DARP is reachable for small instances, up to a hundred requests, using a subsequent amount of computing time. Exact resolution algorithms employ dynamic programming (Desrosiers *et al.*, 1986), branch-and-cut, branch-and-price, and branch-and-price-and-cut. Heuristics can deal with larger instances. Insertion and regret insertion heuristics (Braekers *et al.*, 2014, Diana & Dessouky, 2004), tabu search (Aldaihani & Dessouky, 2003), simulated annealing (Braekers *et al.*, 2014), variable neighborhood search (Parragh *et al.*, 2009), adaptative large neighborhood search (Ropke & Pisinger, 2006) and genetic algorithms (Cubillos *et al.*, 2009) have been proposed in the literature. Cordeau & Laporte (2007) and Ho *et al.* (2018) provide extensive reviews of DARP resolution methods.

In dynamic environments, matching, routing, and rebalancing can also be treated jointly. Exact algorithms designed to solve the static version of DARP can be adapted to deal with dynamic environments where the demand knowledge grows gradually. The most widely used approach consists in solving the static DARP once at the beginning of the planning horizon to obtain a seed solution based on the part of demand known in advance. Then, heuristic methods, such as insertion, deletion, interchange moves, and local search, can be called each time a new request arrives to update the solution (Berbeglia *et al.*, 2010, Luo & Schonfeld, 2011, Vallée *et al.*, 2017). Such a procedure is relevant when a significant number of requests is known in advance and a limited number of requests are to be inserted immediately. This knowledge allows for building good initial routes with sufficient slack to accommodate future immediate requests. The approach fails in highly dynamic environments.

The operator has no clue about the upcoming demand in highly dynamic environments, and online algorithms are necessary. Among online approaches, event-based policies are the most elementary ones. They define specific rules for each event type occurring in the system. Rules can provide the matching policy. For example, the *nearest-idle-vehicle* assigns the nearest currently idle vehicle to the request that has just been emitted in the oversupply regime and assigns the oldest request to the vehicle that has just turned available in the undersupply regime. The *nearest-vehicle* matches the nearest vehicle for which future availability time is known to the oldest open request. The *nearest-idle-vehicle/nearest-open-request* matches the nearest open request with the vehicle that has just turned available in the undersupply regime. Rules can also address ride-sharing as in Levin *et al.* (2017). Whatever the rules defined, the solution is necessarily sub-optimal in the long run, as shown in Figure 4.2

by a simple example borrowed from Psaraftis (1988). Maciejewski (2015) has highlighted the gap between exact offline optimization methods, with full knowledge of demand on the studied period, and event-based heuristics.

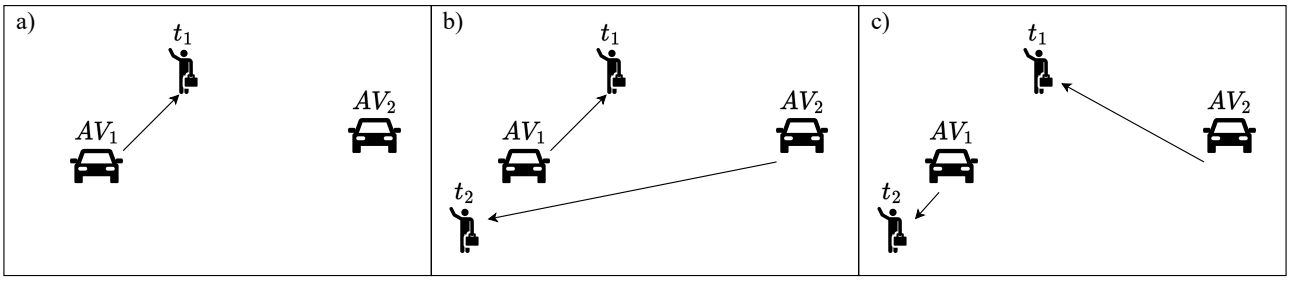


FIGURE 4.2 – *Example of an event-based operational policy leading to a sub-optimal matching decision. (a) Two AVs are available in an area. A traveler sends a request at t_1. AMoD assigns the closest AV, namely AV_1, to this customer. (b) A moment later, another request emerges. If the matching decided previously is permanent, AMoD naturally matches AV_2 with this new request, leading to a sub-optimal matching. (c) The optimal matching is to match AV_1 with t_2 request and AV_2 with t_1 request.*

Two possibilities exist to prevent such sub-optimal decisions. First, one can successively re-optimize matches and routes assigned to the vehicles each time a new request arrives. Jung *et al.* (2016) used simulated annealing combined with insertion heuristics to optimize the entire schedules of vehicles in real-time systematically. Their approach performed better than an event-based policy implementing simple rules. Note that this approach is relevant in the context of AMoD but can pose fairness problems with human drivers since a ride promised to a driver can be canceled at the last minute and assigned to another driver.

Second, the dispatch decision can be delayed by a certain duration. In this way, dispatch decisions are made on a batch of requests. It allows managing the fleet in a less myopic way than in event-based policies. Batch-matching allows reaching lower waiting times than event-based policies, as shown by Maciejewski *et al.* (2016). Uber currently uses it. Hyland & Mahmassani (2018) test six batch-matching strategies in a highly dynamic environment over a Manhattan network and compare them to event-based rules. Alonso-Mora *et al.* (2017b) relies on batch-matching but the method proposed goes further by including routing and ride-sharing. Ke *et al.* (2019) proposes a Multi-Agent Reinforcement Learning (MARL) framework to decide the delay each request should undergo before joining the batch. Compared to matching without delays, the learned policy reduces the average pick-up time with little loss on the order response rate.

Rebalancing is an external module for event-based, systematic re-optimization, and batch-matching approaches. As AVs can continuously rebalance contrary to the traditional car or bike sharing systems where rebalancing can be achieved at most a few times in a day, AMoD-specific tools have been proposed in the literature, such as linear programs (Alonso-Mora *et al.*, 2017b, Zhang *et al.*, 2019), model predictive control (Carron *et al.*, 2021), or reinforcement learning (Lin *et al.*, 2018, Yu *et al.*, 2019, Zhang *et al.*, 2020, Liu *et al.*, 2020).

Reinforcement Learning (RL) has been used for rebalancing only and for matching and rebalancing jointly in a centralized (Jin *et al.*, 2019, 2020, Mao *et al.*, 2020) or decentralized manner (Gueriau *et al.*, 2020). Learned operational policies are becoming popular in the

literature.

4.3.1.3 Characteristics of AMoD environment

The choice for an approach to deal with AMoD operational policy in our pricing scheme optimization framework relates to the characteristics of AMoD and the environment in which we consider it.

First, AMoD demand cannot be known in advance for the whole studied period. Second, we want to tackle a large urban area case with many AVs and requests. The operational policy algorithm should be efficient in computing time and scalable. Due to these points, exact resolution approaches for the DARP are not adapted.

Third, we want to tackle pricing scheme optimization. Learned policies are not amenable to this objective. An operational policy is learned under a specific pricing scheme. It is not necessarily relevant under another scheme. The only way is to jointly learn the regulator pricing scheme and the AMoD operational policy with a MARL framework. Such an approach is complex and has recently shown results for small-scale and simplified scenarios only (Shou & Di, 2020).

Fourth, we do not enable ride-sharing in this thesis. We focus on improving the transportation system's efficiency by fostering cooperation between autonomous ride-hailing and PT. Rides are shared only on the PT legs of trips. Enabling ride-sharing for the AMoD is another way to reduce VKT and reach a greener operation for the whole system. This option is out of the scope of this thesis but is a relevant extension for the work presented here.

Fifth, we are not seeking optimality here. The policy should reflect the profit-oriented behavior of AMoD operation as best as possible in a highly dynamic environment.

Six, we are interested in the morning commute in an urban context where mobility patterns are usually asymmetric. The AMoD system should be able to serve a far away profitable request with a reasonable pick-up time for the customer. Combining an event-based, systematic re-optimization, or batch-matching with an external rebalancing module is a solution. While the matching policy operates on the requests currently open, rebalancing operates on future requests. Another solution is to run batch-matching on a horizon. It simplifies the policy since no external rebalancing module is required. Then, the fleet's behavior comes down to a matching and routing problem. This solution is chosen here. The rest of the chapter describes and analyzes batch-matching over horizon policy. Notably, we wonder if such a simple approach is efficient in correctly managing the fleet, i.e., maximizing the AMoD operator profit while being sensitive to regulator taxes and subsidies.

4.3.2 Description of the batch-matching over horizon policy

4.3.2.1 Problem setting

Let t be time and H the horizon length. At t, the AMoD operator knows all requests already issued and all requests that will be issued between t and $t + H$. The set of known requests at t is R, and the set of known open requests at t is R_O. A request is *open* when it has not been permanently matched and has not been canceled by the traveler ($R_O \subset R$). Each request

$r \in R$ has an earliest pick-up time e_r (which is equal to the request time in our case) and a latest pick-up time l_r. A maximum waiting time w_r is defined for each traveler ($l_r = e_r + w_r$).

As we focus on ride-hailing, the capacity of all AVs is one. We call V the set of all AVs composing the fleet and V_O the set of available AVs at t. An AV is *available* when it has no more than $K \geq 0$ requests in its service plan. We note R_v the set of requests that have been permanently matched with $v \in V$ and not yet served. The service plan of v corresponds to the ordered set of requests in R_v. It is noted $P_{R_v} = \{r_1, ..., r_k\}$, where $k \leq K$.

AMoD's operational policy matches the known requests with the available AVs, updates vehicles' service plans, and eventually orders idle AVs to reposition. The policy should maximize the total profit earned by the fleet during a given working period. The following sections describe the batch-matching over horizon policy principles: ***permanent*** and ***temporary*** modes for the rolling horizon, ***impatient*** and ***productivist*** definitions for the utility of a match, and optimal matching. Section 4.3.2.5 details the associated algorithm.

4.3.2.2 Permanent and temporary modes

Figure 4.3 presents the rolling horizon principle used in the algorithm. We implement two modes in the algorithm: a permanent matching mode (*permanent*) and a short-term permanent matching / long-term temporary repositioning mode (*temporary*). For both modes, let α be a float in $[0, 1]$. In *permanent* mode, all matches identified by the optimal matching resolution are made permanent. In *temporary* mode, only matches (v, r) implying a request r satisfying $e_r \leq t + \alpha H$ are made permanent. Other matches (v, r) identified are so that $e_r > t + \alpha H$. They lead to the initiation of repositioning movements by the concerned AVs. The temporary match is not registered in AV's service plan as it is not a service mission. At the next optimal matching resolution, repositioning AVs are therefore considered available. Note that repositioning movements are only accessible to idle AVs, i.e., those having no mission in their service plan. The *temporary* mode increases AMoD flexibility by putting back into play a part of the decisions taken. In both modes, the horizon is rolling, i.e., the optimal matching resolution is called each αH.

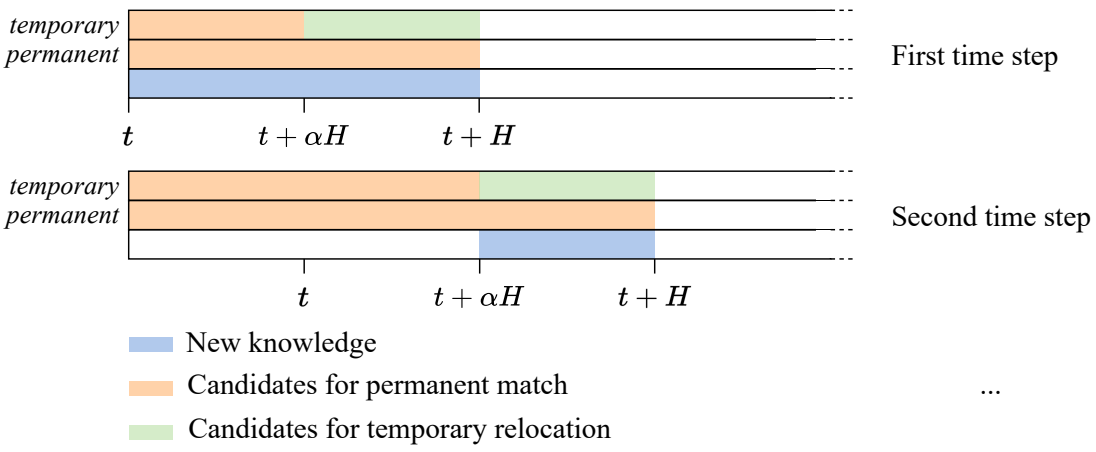


FIGURE 4.3 – *Rolling horizon principle for permanent and temporary modes.*

4.3.2.3 Impatient and productivist utilities

A request $r \in R$ is characterized by a gross income p_r, which is the price traveler pays to the AV. As stated in section 4.1, AMoD has no long-term economic strategy. It has a unique fixed fare scheme over the urban area. This fare scheme is competitive and ensures lucrative rides for AVs. It is more attractive economically for travelers than today's (e-)ride-hailing services. Here, we use one of the most popular fare schemes in literature (Gurumurthy *et al.*, 2020, 2021, Wen *et al.*, 2018, Simoni *et al.*, 2019). Values for the base, distance-based, and time-based fares are chosen to overcome the production cost (Becker *et al.*, 2020) of any ride an AV performs. AMoD profit is positive for any isolated ride (without including the relocation cost). The AMoD fare scheme is:

$$p_r = f_0^{\mathrm{a}} + f_{\mathrm{d}}^{\mathrm{a}} d_r + f_{\mathrm{t}}^{\mathrm{a}} t_r \tag{4.1}$$

where f_0^{a} is the base fare, $f_{\mathrm{d}}^{\mathrm{a}}$ is the distance-based fare, $f_{\mathrm{t}}^{\mathrm{a}}$ is the time-based fare, d_r is the travel distance of the path traveler wants to ride onboard AV, t_r the travel time on this path.

The pair (v, r) designates a potential match between v and r. It is characterized by a utility $u_{v,r}$ defined as follows:

$$u_{v,r} = u(P^*_{R_v \cup \{r\}}) - u(P_{R_v}) \tag{4.2}$$

where $u(P_{R_v})$ designates the utility of the current service plan of v, $u(P^*_{R_v \cup \{r\}})$ is the utility of the best plan including all requests in R_v and request r. By best plan, we mean the plan with the highest utility that satisfies the time windows of all requests in R_v.

We define two strategies AMoD can adopt to maximize its profit. The *impatient* strategy (equation 4.3) takes into account the repositioning cost of v toward the pick-up points of requests in the plan, the expected income for serving requests in the plan, and the eventual time lost by v while waiting for traveler departure (at e_r). The waiting time AV experience by arriving early at a pick-up point is penalized. The *productivist* strategy (equation 4.4) takes into account the repositioning cost and the expected income into a profit per time unit, i.e., a profit earned in terms of time rate. Earning more profit in less time is favored. These two strategies lead to two definitions of the utility of a service plan.

$$u_{\mathrm{imp}}(P_{R_v}) = \begin{cases} 0, & \text{if } |R_v| = 0 \\ \sum_{i=1}^{k} -C_{\mathrm{d}}^{\mathrm{a}} d_{r_{i-1},r_i} + (p_{r_i} - C_{\mathrm{d}}^{\mathrm{a}} d_{r_i}) \xi_{r_i|P_{R_v}} - f_{\mathrm{t}}^{\mathrm{a}} \max(0, e_{r_i} - E_{r_i|P_{R_v}}), & \text{if } r_1 \text{ unpicked} \\ \sum_{i=2}^{k} -C_{\mathrm{d}}^{\mathrm{a}} d_{r_{i-1},r_i} + (p_{r_i} - C_{\mathrm{d}}^{\mathrm{a}} d_{r_i}) \xi_{r_i|P_{R_v}} - f_{\mathrm{t}}^{\mathrm{a}} \max(0, e_{r_i} - E_{r_i|P_{R_v}}), & \text{otherwise} \end{cases} \tag{4.3}$$

$$u_{\text{prod}}(P_{R_v}) = \begin{cases} 0, & \text{if } |R_v| = 0 \\ \dfrac{\sum_{i=1}^{k} -C_{\text{d}}^{\text{a}} d_{r_{i-1},r_i} + (p_{r_i} - C_{\text{d}}^{\text{a}} d_{r_i}) \xi_{r_i|P_{R_v}}}{L_{r_k|P_{R_v}} + t_{r_k} - t}, & \text{if } r_1 \text{ unpicked} \\ \dfrac{\sum_{i=2}^{k} -C_{\text{d}}^{\text{a}} d_{r_{i-1},r_i} + (p_{r_i} - C_{\text{d}}^{\text{a}} d_{r_i}) \xi_{r_i|P_{R_v}}}{L_{r_k|P_{R_v}} + t_{r_k} - (L_{r_1|P_{R_v}} + t_{r_1})}, & \text{otherwise} \end{cases} \tag{4.4}$$

where C_{d}^{a} designates the distance-based operation cost of an AV, d_{r_{i-1},r_i} ($i \geq 2$) is the distance of the shortest path between r_{i-1} drop-off point and r_i pick-up point, d_{r_0,r_1} is the distance between v's current location and r_1 pick-up point, $E_{r_i|P_{R_v}}$ is the time at which v arrives at r_i pick-up point following plan P_{R_v}, $L_{r_i|P_{R_v}}$ is the time at which v leaves r_i pick-up point following plan P_{R_v}, $\xi_{r_i|P_{R_v}}$ is a binary variable which equals 1 if $E_{r_i|P_{R_v}} \leq l_{r_i}$ (r_i time window is verified for P_{R_v}) and 0 otherwise (r_i time window is not verified for P_{R_v}). AMoD is assumed to know l_r and e_r for each request $r \in R$. Equations 4.5 to 4.7 specify how $E_{r_i|P_{R_v}}$ and $L_{r_i|P_{R_v}}$ can be recursively computed.

$$E_{r_1|P_{R_v}} = \begin{cases} t + t_{r_0,r_1}, & \text{if } r_1 \text{ unpicked} \\ t - t_{r_1,r_0}, & \text{otherwise} \end{cases} \tag{4.5}$$

where t_{r_0,r_1} is the travel time on the shortest path between v current location and r_1 pick-up point, t_{r_1,r_0} is the travel time on the shortest path between r_1 pick-up point and v current location.

$$L_{r_i|P_{R_v}} = \max(e_{r_i}, E_{r_i|P_{R_v}}) \tag{4.6}$$

$$E_{r_{i+1}|P_{R_v}} = L_{r_i|P_{R_v}} + t_{r_i} + t_{r_i,r_{i+1}} \tag{4.7}$$

where $t_{r_i,r_{i+1}}$ is the travel time on the shortest path between r_i drop-off point and r_{i+1} pick-up point.

We test two strategies to compute $P^*_{R_v \cup \{r\}}$. In the first one, we only envisage one potential plan where the new request is inserted at the end of the plan. In the second, we try to insert r at each index of P_{R_v} and keep the one with the maximal utility as a potential plan. The first index is not tried when r_1 has already been picked up. Testing all possible permutations of $R_v \cup \{r\}$ is unnecessary here. As the AV plan is incrementally built by inserting at most one request per call of the optimal matching resolution, using an insertion heuristic is sufficient. For both strategies, the incremented plan is a real candidate only if its utility is positive ($u(P^*_{R_v \cup \{r\}}) \geq 0$) and if the time windows of all requests in R_v are satisfied ($\xi_{r_i|P_{R_v \cup \{r\}}} = 1$ for $r_i \in R_v$).

Note that for $r \in R_{\text{O}}$ with $e_r > t + \alpha H$ and $v \in V_{\text{O}}$ with $|R_v| > 0$, we set $u_{v,r}$ to a negative float so that only idle AVs are candidates for repositioning movements.

4.3.2.4 Optimal matching

Based on the utilities of all potential matches, the following optimization problem is solved:

$$\max_{x_{v,r}} \quad \sum_{v \in V_O} \sum_{r \in R_O} u_{v,r} x_{v,r} \tag{4.8a}$$

$$\text{subject to} \quad x_{v,r} \in \{0,1\}, \forall v \in V_O, \forall r \in R_O \tag{4.8b}$$

$$\sum_{v \in V_O} x_{v,r} \leq 1, \forall r \in R_O \tag{4.8c}$$

$$\sum_{r \in R_O} x_{v,r} \leq 1, \forall v \in V_O \tag{4.8d}$$

where $x_{v,r}$ are the binary decision variables, equal to 1 if (v, r) match is decided, 0 otherwise.

To prevent unfavorable interactions between two matches for one AV decided during the same dispatch round, constraint 4.8d imposes that an AV can be assigned to at most one request per optimal matching resolution call. Indeed, when several matches are allowed for one AV v, utility $u_{v,r}$, $r \in R_O$, is directly dependent on the assignment of other requests to the same AV. Separately, matches (v,r) and (v, r') can have high utilities, but it does not mean that utility of a route including requests r and r' has a good utility.

Due to constraint 4.8d, there is a risk for under-exploited AVs. Some AVs may finish their mission before the next call of the optimal matching problem and remain idle for a few minutes. To prevent this phenomenon, we adjust α and introduce several successive service missions in AV planning (at most K, where missions are appended in different calls of the optimal matching resolution). The value of α should not be too small to prevent calling the optimal matching resolution too often and not too big to limit the number of lost resources between two calls. It should be consistent with the average riding time of a service mission in the considered scenario.

Finally, constraint 4.8c ensures that a request is matched with at most one AV.

A state-of-the-art solver implementing a branch & cut algorithm is used in this chapter to solve the optimal matching problem.

4.3.2.5 Algorithm

Batch-matching over horizon algorithm is detailed below.

Algorithm 2: Batch-matching over horizon algorithm

1 Initialize V_O to V, R and R_O to empty lists, t to 0;
2 Set horizon to $[t, t+H]$, gather requests r emitted within the horizon and append new ones to R and R_O, gather AVs v with at most K requests in their service list to form V_O;
3 Calculate utilities $u_{v,r}$ for each (v, r) pair, $v \in V_O$, $r \in R_O$;
4 Form a bipartite graph with $v \in V_O$ and $r \in R_O$ as nodes, (v, r) as arcs and $u_{v,r}$ as cost on arcs. Compute the matching for which the sum of utilities is maximal where one vehicle is assigned to at most one new request.;
5 *permanent*: Make permanent all matches found in step 4. *temporary*: Make permanent all matches found in step 4 for which $e_r \in [t, t+\alpha H]$, and initiate temporary repositioning movements for the remaining matches.;
6 Remove permanently matched requests from R_O;
7 Set $t = t + \alpha H$ and go to step 2;

4.4 Analysis of the batch-matching over horizon approach

4.4.1 Testing environment

We want to analyze the batch-matching over horizon policy in conditions close to the ones in which we want to optimize the zone-based pricing scheme. To do so, we developed a simple agent-based simulation platform. The simulator includes the components to model AMoD interactions with travelers within a multimodal transportation network. Notably, it renders travelers' objective to minimize their travel cost, AMoD's objective to maximize its profit through the batch-matching over horizon policy, and defines the rules under which AMoD and travelers interact. A theoretical urban area with a severe asymmetry in demand pattern is implemented. This unfavorable scenario for AMoD allows for challenging the operational policy. This section describes the simulation platform developed and the instance used to test and analyze the batch-matching over horizon policy.

4.4.1.1 Multimodal network

To represent the different transportation modes, including walk (w), car (c), AV (a), train (r), subway (s), and bus (b), a digraph $G = (N, A)$, as the one shown in Figure 4.4, is used. N and A are the sets of vertices and arcs. Each mode has an associated layer in this graph. Thus, $G_w = (N_w, A_w)$ is the walking layer, $G_c = (N_c, A_c)$ is the personal car layer, $G_a = (N_a, A_a)$ is the AV layer, $G_r = (N_r, A_r)$ is the train layer, $G_s = (N_s, A_s)$ is the subway layer and $G_b = (N_b, A_b)$ is the bus layer. The walk layer is based on walkable streets. The car and AV layers are based on the road network: vertices and arcs correspond to road intersections and links. The train, subway, and bus layers are based on the public transportation network: vertices and arcs correspond to transit stations and itineraries between two stations. A set of transfer arcs connect the walk layer to the other layers. Transfer arcs types are: *starter* (walk $\rightarrow$ car), *park* (car $\rightarrow$ walk), *pick-up* (walk $\rightarrow$ Av), *drop-off* (AV $\rightarrow$ walk), *board* (walk $\rightarrow$ train, walk $\rightarrow$ subway, walk $\rightarrow$ bus), *alight* (train $\rightarrow$ walk, subway $\rightarrow$ walk, bus $\rightarrow$ walk). Another

set of transfer arcs of type *alighboard* are internal to the urban public transportation network and connect the subway and bus layers.

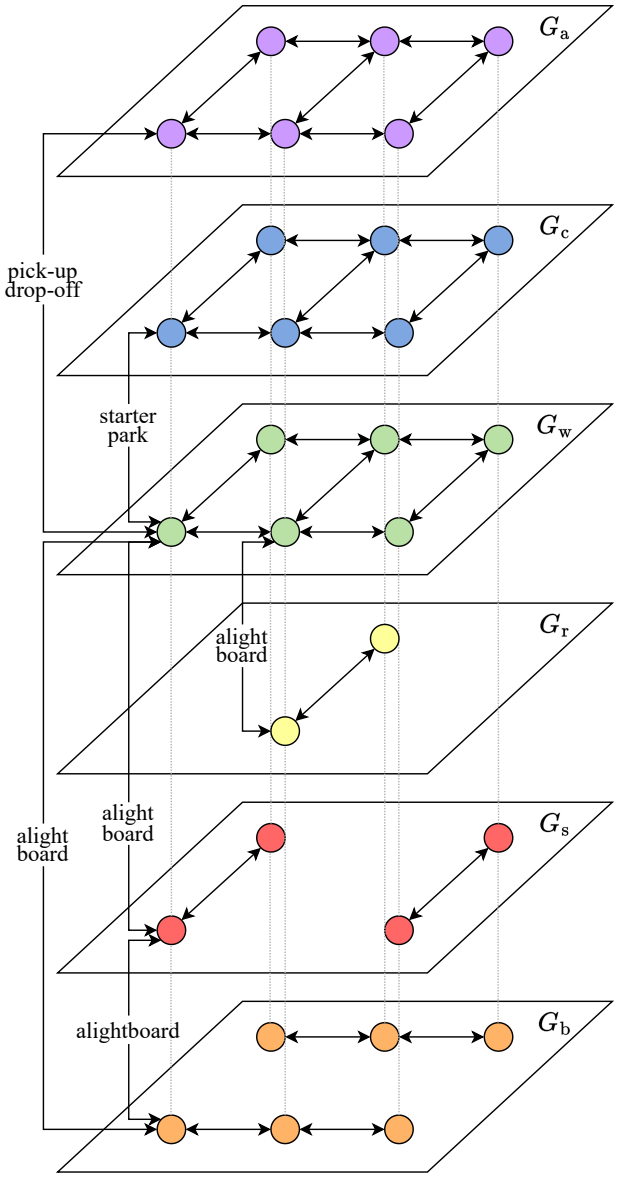


FIGURE 4.4 – *The multimodal network in our agent-based simulation platform comprises a walk, car, AV, train, subway, and bus layer. Travelers can transit between two layers thanks to transfer arcs that connect the walk layer to other layers. As the subway and buses belong to the same public transportation operator, alightboard transfer arcs link both layers.*

4.4.1.2 Travel times

Intra-layer arcs belonging to A_c and A_a are parameterized with a mean speed. A fixed walking speed v_w applies on all walking arcs. Travel times on all arcs in $A_c \cup A_a \cup A_w$ are deduced from speed and distance. Each transit line is characterized by a cruising speed (v_r, v_s, v_b), a headway (h_r, h_s, h_b) and a time lost per station (τ_r, τ_s, τ_b). The travel time between two stations of the same transit line is the sum of time lost per stop and travel

time at cruising speed. The travel time between two stations of different transit lines located at the same coordinates is approximated to be half the headway, such as the travel time between a walk node and a transit station with the same coordinates. The travel times on *starter*, *park*, *drop-off*, *alight* arcs are considered null. Travel times on *pick-up* arcs depend on the level of service provided by AMoD in practice, but a reference pick-up time (τ_a) is used for travelers' itinerary choices.

4.4.1.3 Travel costs

The itinerary choice is considered deterministic and based on generalized travel costs. Travel cost includes a monetary cost, a travel time cost, and a transfer penalty term. The monetary cost associated with a car itinerary includes a distance-based cost (C_d^c), carried by the *car* arcs, and a parking cost carried by the *park* arcs. On each *park* arcs, a certain parking capacity (K^c) and a price (p^c) are defined.

The train pricing scheme comprises a base fare (f_0^r) and a distance-based fare (f_d^r). The urban public transport network pricing scheme only includes a base fare ($f_0^{s,b}$) that should be paid once at the network's entrance. The AMoD pricing scheme has been presented in section 4.3.2.3.

To homogenize time and money, we use several values of time depending on the activity (β_{ride}, β_{drive}, β_{wait}). In such an intermodal network, the transfer cost cannot be ignored Gallotti & Barthelemy (2015). The disutility associated with transferring from one mode to another has largely been studied, particularly in transit systems Currie (2005), Garcia-Martinez *et al.* (2018). This is taken into account through a monetized transfer penalty $\gamma_{transfer}$. An adapted version of the Dijkstra algorithm is used to account for transfer penalties while exploring the graph (see appendix B.1 for more details).

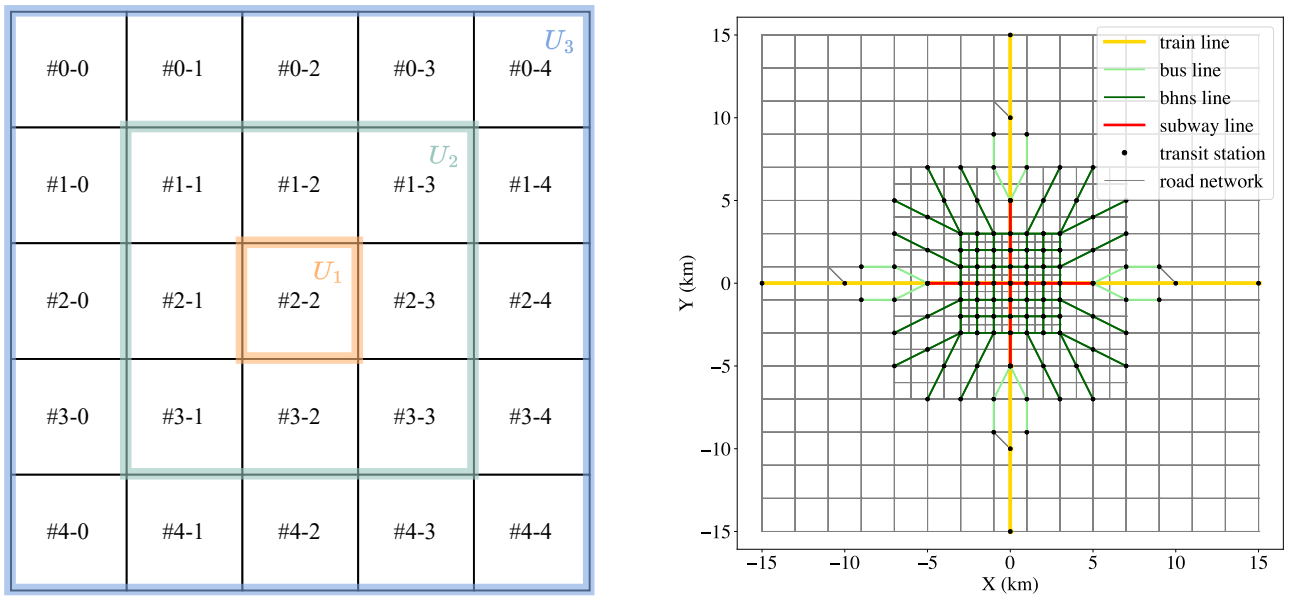
4.4.1.4 A scenario with tough demand-supply imbalances

The simulator takes a road network as an input. In our theoretical case, we chose a Manhattan road network with three mesh sizes to account for topology differences over the urban area. The smaller mesh size stands in the center of the city (U_1), the medium one in the suburbs (U_2), the biggest one in the extended suburbs and the close rural area surrounding the city (U_3). Figure 4.5a presents this zoning. Two ring roads are added on the boundaries between U_1-U_2 (U_{12}) and U_2-U_3 (U_{23}). A fixed mean speed (v_{U_k}) for cars and AVs is defined in each urban zone. Vehicles travel arcs of the network node by node at mean speed. Traffic dynamics are not modeled here. In addition, since the number of AVs circulating remains limited compared to background traffic, we assume they have no impact on the known mean speed. Note that the simulation platform used in chapter 5 includes a traffic model based on the trip-based MFD framework (Mariotte, 2018) to account for traffic dynamics.

A hybrid transit system, as the one proposed by Daganzo (2010), Estrada *et al.* (2011), is chosen for the monocentric city (Figure 4.5b).

The morning commute demand pattern is a many-to-one: all travelers' destinations are in the center. Regarding network topology and travel costs, most of them prefer to request AV for a door-to-door ride. A marginal part of them requests AV for the first mile to join a

transit station. The area splits into 25 service zones from #0-0 to #4-4 (Figure 4.5a). Figure 4.6 shows the demand-supply imbalance. In total, 4000 requests are emitted during 3 hours.



(a) *Urban (in colors) and service zoning of the studied area.*

(b) *The multimodal network instance.*

FIGURE 4.5 – *A theoretical urban area.*

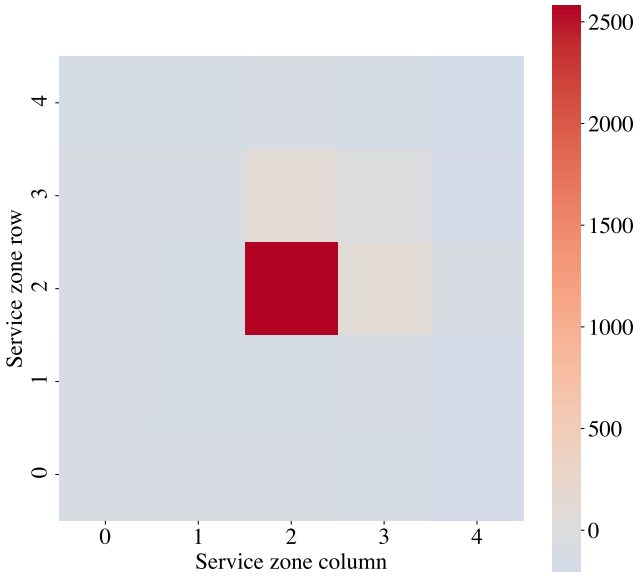


FIGURE 4.6 – *Number of pick-ups minus number of drop-offs requested per zone. A many-to-one demand pattern is considered in the theoretical instance to provoke a tough imbalance between supply and demand.*

This scenario has tough demand-supply imbalances and requires repositioning from the AMoD operator. Indeed, the latter manages a limited number of AVs within a large coverage zone during the morning commute in a monocentric city. Consequent repositioning efforts

are required by AVs, especially for distant requests from the center. The operator does have choices to make between requests. Some will be prioritized to the detriment of others. We chose this scenario because it challenges the AMoD operator and allows for highlighting the impact of the operational policy on this prioritization.

All parameters can be found in table 4.2.

Road network			
v_{U_1}	0.55	(m/s)	Mean speed in U_1
v_{U_2}	0.67	(m/s)	Mean speed in U_2
v_{U_3}	1.17	(m/s)	Mean speed in U_3
$v_{U_{12}}$	1.09	(m/s)	Mean speed on U_{12} ring road
$v_{U_{23}}$	1.34	(m/s)	Mean speed on U_{23} ring road
$C_{\mathrm{d}}^{\mathrm{c}}$	0.6	(EUR/km)	Operation cost of personal car
ρ_{U_1}	50	%	Probability of free private parking in U_1
$p_{U_1}^{\mathrm{c}}$	6	(EUR)	Parking price in U_1
$K_{U_1}^{\mathrm{c}}$	10	(parking spots)	Number of parking spots per *park* edge in U_1
Transit			
v_{r}	1.34	(m/s)	Cruising speed of train
τ_{r}	2	(min)	Time lost per station for train
h_{r}	15	(min)	Headway of train lines
f_0^{r}	0.5	(EUR)	Base fare for train
$f_{\mathrm{d}}^{\mathrm{r}}$	0.2	(EUR/km)	Distance-based fare for train
v_{s}	0.9	(m/s)	Cruising speed of subway
τ_{s}	1	(min)	Time lost per station for subway
h_{s}	4	(min)	Headway of subway lines
v_{b}	0.55	(m/s)	Cruising speed of bus
τ_{s}	1.5	(min)	Time lost per station for bus
h_{s}	6	(min)	Headway of bus lines
$f_0^{\mathrm{b,s}}$	1.5	(EUR)	Fare for entering transit network
v_{w}	0.11	(m/s)	Walking speed
AMoD			
f_0^{a}	1.3	(EUR)	Base fare for AMoD
$f_{\mathrm{d}}^{\mathrm{a}}$	0.3	(EUR/km)	Distance-based fare for AMoD
$f_{\mathrm{t}}^{\mathrm{a}}$	0.3	(EUR/min)	Time-based fare for AMoD
$C_{\mathrm{d}}^{\mathrm{a}}$	0.27	(EUR/km)	Operation cost of AV
τ_{a}	5	(min)	Reference pick-up time
Travelers			
β_{wait}	0.425	(EUR/min)	Value of time when waiting
β_{ride}	0.17	(EUR/min)	Value of time when riding transit of AV
β_{drive}	0.255	(EUR/min)	Value of time when driving
$\gamma_{\mathrm{transfer}}$	0.85	(EUR)	Transfer penalty

TABLE 4.2 – *Theoretical instance parameters.*

4.4.1.5 AMoD-travelers interactions

AMoD and travelers interact within this environment following the scheme 4.7. A traveler who chooses an itinerary including an AV ride sends her request r to AMoD. It contains timing constraints computed with w_r. The dispatcher knows all requests on the horizon and receives AVs' states, positions, and service plans in real-time. It resolves the optimal matching problem and decides to match r with v. For *permanent* strategy, r is immediately added to v's service plan. For *temporary*, if e_r belongs to the second part of the horizon and v service plan is empty, v only receives the order to reposition toward r's pick-up point till the next matching round. For a permanent match, r is picked-up by v as soon as v arrives, except if e_r is not yet reached. In this case, v waits for r. Then, v picks r up, carries him to his destination, and drops him off. For non-AV legs of the itinerary, the customer travels node by node on the other graph layers. Travelers who have chosen car mode book parking in advance to prevent overcoming parking capacities. Behaviors for AVs and travelers are detailed in appendices B.2 and B.3.

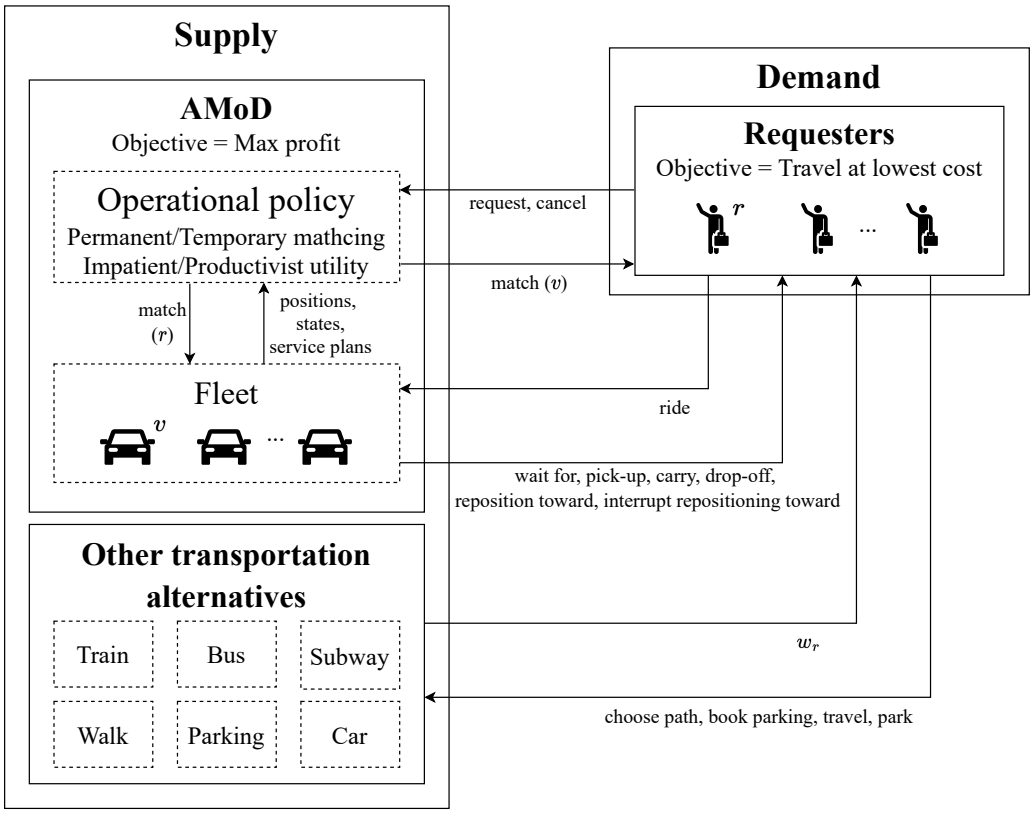


FIGURE 4.7 – *AMoD-travelers interactions.*

4.4.1.6 Implementation choices

Note that for the objectives of chapter 4, the explicit modeling of other transportation alternatives than AMoD is not mandatory, especially since travel times and pricing schemes are considered static in this chapter. In practice, we could have launched a simulation including only the AMoD module with a set of requests extracted from the global demand, and the travel times/distances between each node of the multimodal network as parameters.

We decided to keep the global demand and the multimodal network in the simulation platform and pre-computed the "shortest" paths (in generalized cost for travelers, in distance for AVs) between each node. In this way, the simulation time is not affected, and we have a generic simulation platform that can be enriched and used for studying the effect of online regulation measures, such as a dynamic pricing scheme, in the future. The platform could be incremented with a PT module (actual circulation of PT rolling stocks within the network) and a congestion dynamics module to address such problems.

4.4.2 Numerical results

The horizon length H is a crucial parameter for batch-matching over horizon policy. In this section, we present the results of the sensitivity of H on AMoD-side and travelers-side indicators. We highlight the policy variant and the horizon length allowing the operator to manage the fleet best for the scenario presented above. For this variant, we check how the batch-matching over horizon policy performs compared to an event-based heuristic. Finally, we check the sensitivity of the fleet's behavior to a naive zone-based pricing scheme.

4.4.2.1 Investigated indicators

Six indicators are tracked.

- AMoD-side indicators are:
  - **P**, total profit earned during the whole studied period
  - **TDTS**, total distance traveled by serving AVs
  - **TDTE**, total distance traveled by empty AVs, including AVs that are permanently matched and AVs that are temporarily repositioning
- Travelers-side indicators are:
  - **ORR**, the order response rate, defined as the ratio between the number of matched requests and the total number of requests emitted
  - **TWT**, total waiting time for AVs, including the total realized pick-up time, the total waiting time of travelers that have canceled, and the total waiting time of travelers that are still waiting at the end of the studied period
  - **G**, the Gini coefficient of zones' ORRs, reflects the inequality of ORR among service zones (the higher is G, the more unequal the system)

4.4.2.2 Choice for α, K and insertion heuristic

To better highlight the differences between the proposed strategies, we let more freedom to the dispatcher by setting the maximum waiting time of all travelers to a substantial value $W = 20$min. The fleet size is 500 AVs.

As discussed in section 4.3.2, proper values for α and K should be chosen. On the one hand, increasing α may reduce the computing time since the optimal matching resolution

is called fewer times. On the other hand, decreasing α may help increase the ORR by overcoming the "at most one match per AV per matching round" constraint. Similarly, decreasing K may reduce the computing time since the size of R_O at each matching round is smaller. Increasing K may increase the ORR by exploiting each AV more.

We launched the sensitivity analysis of H on the defined indicators for several parameters. We compared αH=5, 10min, K=1, 3, under the two insertion heuristics proposed in 4.3.2.3 (insert request at the end of the plan, or insert it at the best index of the plan). Comparing αH=5 with αH=10 shows that α has a marginal impact on fleet behavior. P and ORR profiles are quite similar, while the computing time gain increases with H when the matching problem is called twice less, as shown in Figure 4.8a. The gain is higher for *temporary* strategies. For permanent strategies with K=1, the gain remains null for all H values. These observations are valid for K=1 and K=3 under both insertion heuristics. Then, **we choose αH=10min**.

Comparing K=1 with K=3 shows that K has a marginal impact on P and ORR for the *temporary* strategies. The *permanent-impatient* strategy has an increasing ORR from H=20min when K=3 while it has a decreasing ORR from the same horizon length when K=1 (Figure 4.8b). For a higher ORR, it has a slightly lower P. Indeed, for a fair comparison, P is the total profit realized during the fixed studied period. At the end of this period, many permanent matches are not yet realized by the AVs. These observations are valid for both insertion heuristics. As P is not tremendously improved by increasing K under our demand pattern, **we decide to choose K=1**.

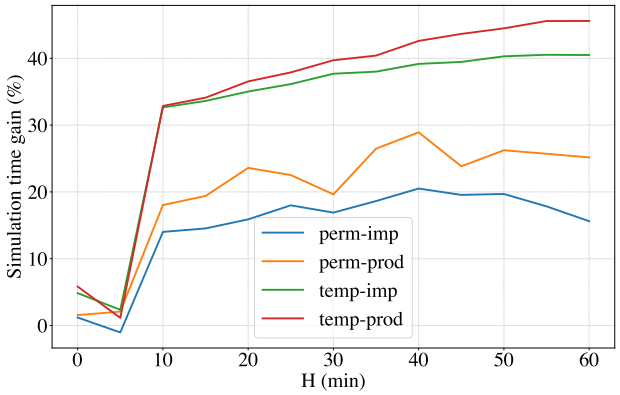
Comparing both insertion heuristics shows that the "insert at best plan index" heuristic improves the ORR for all H under the *permanent* strategies (Figure 4.8c). It improves the ORR of the *temporary* strategies more slightly. The profit curves, however, are similar for both heuristics (Figure 4.8d). The computation time gain with the "insert at plan end" heuristic varies from 0 to 45% depending on the variant and H. Then, **we choose to use the "insert at plan end" heuristic**.

4.4.2.3 Sensitivity of H on indicators

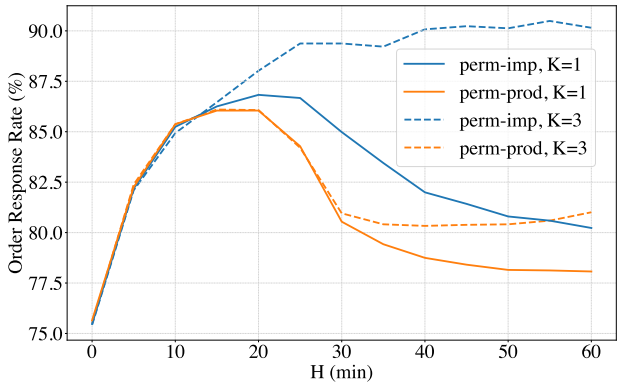
With the chosen parameters, our four matching strategies are compared on their sensitivity to H. Figure 4.9 presents the results.

Batch-matching over horizon enables AMoD to serve more requests whatever the strategy compared to traditional batch-matching (H=0min) ORR starts by increasing with H. The fleet is underused at H =0min, with at least a hundred AVs being idle at any time. The lack of horizon prevents AVs from being dispatched to distant requests. The repositioning time required to join pick-up points in corner zones (#0-0, #0-4, #4-0, #4-4) from #2-2 (where most AVs finish their service mission), is higher than W. Consequently, inequalities in ORR among zones are significant, as shown in Figure 4.10a and confirmed by G, above 0.2 for all strategies. A horizon of 15-25min maximizes the number of matches. The maximal value of ORR reached is between 85% and 87%, depending on the strategy.

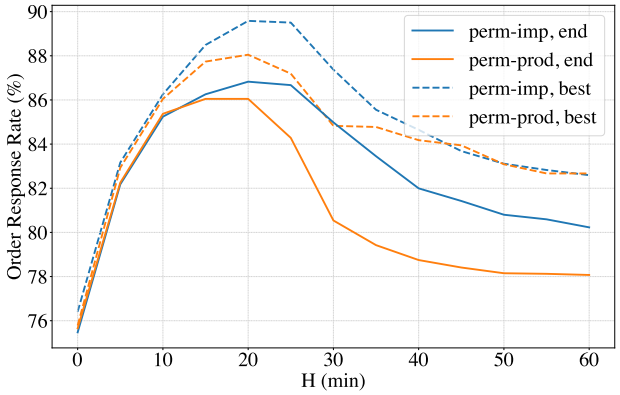
The *impatient* utilities are better to maximize ORR and minimize G, to the detriment of P The *impatient* utilities perform respectively better than their counterparts to maximize



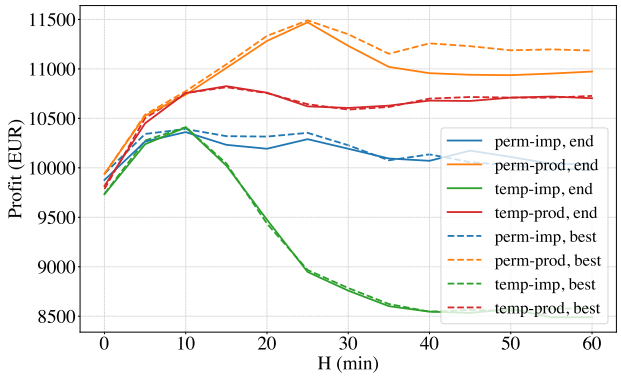
(a) *Simulation time gain between αH=5min and αH=10min for K=3 with the "insert at the end of plan" heuristic.*



(b) *Comparison of ORR between K=1 and K=3 for αH=10min with the "insert at the end of plan" heuristic.*



(c) *Comparison of ORR between insertion heuristics for αH=10min and k=1.*

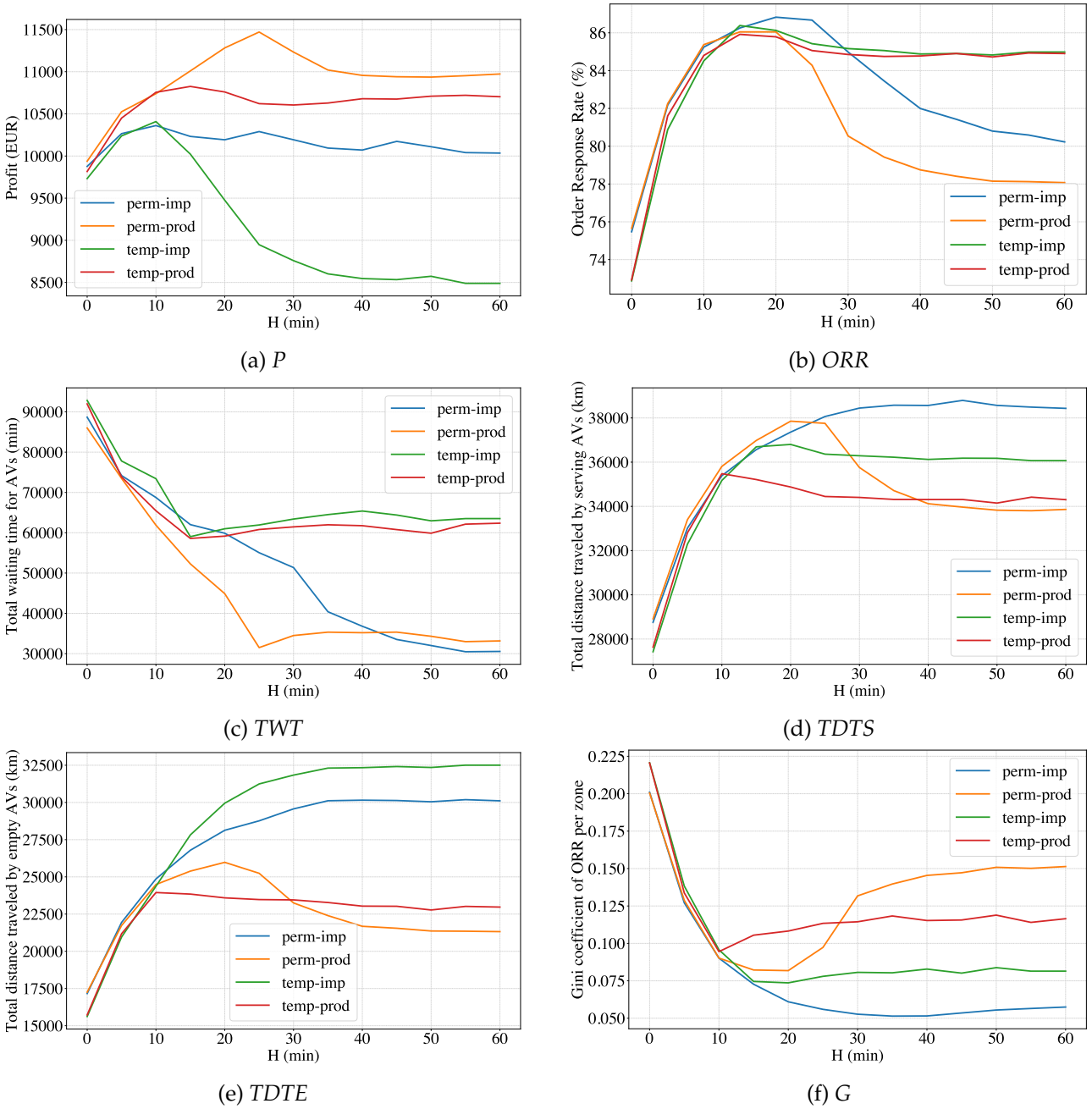


(d) *Comparison of P between insertion heuristics for αH=10min and k=1.*

FIGURE 4.8 – *Choice of α, K, and insertion heuristic.*

ORR and minimize G. Each erases more ORR disparities than a *productivist* utility under the same horizon management mode. Figures 4.10b and 4.10c represent the ORR per zone under *permanent-impatient* and *temporary-productivist*, that are the strategies respectively minimizing and maximizing G for H=20min. With the same knowledge of future requests, the raise of ORR in U_3 corner zones is better for *permanent-impatient*. However, the *impatient* strategies are respectively behind their counterparts in maximizing P.

Two distinct behaviors emerge from the two utility definitions For *permanent-impatient*, distant requests for which AVs can arrive within the earliest pickup time - latest pickup time range are more attractive than closer requests for which AVs would have to wait for the customer. For *permanent-productivist*, AVs are matched in priority with closer requests even if they must wait for the customer. AV waiting time is included in the *productivist* utility definition, contributing to the time required to earn a given profit. However, it does not have as much importance as in the *impatient* utility. The average number of waiting AVs for H=20min in *permanent-productivist* has a higher score (35 AVs) than in the other three (between 4 and 15 AVs). For *temporary-productivist*, we do not find a consequent number since only AVs waiting for a customer involved in a permanent match are counted.

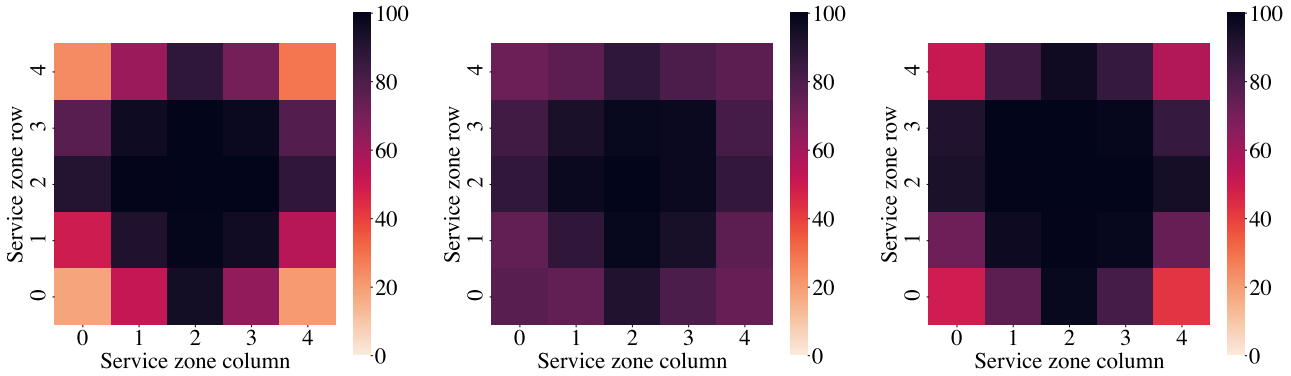


(a) *P* (b) *ORR*

(c) *TWT* (d) *TDTS*

(e) *TDTE* (f) *G*

FIGURE 4.9 – *Sensitivity of H on indicators for the four variants of the batch-matching over horizon.*



(a) *H=0min, permanent-impatient* (b) *H=20min, permanent-impatient* (c) *H=20min, temporary-productivist*

FIGURE 4.10 – *ORR per zone for different variant and H.*

Extending horizon length more than 15-25min has almost no effect on ORR for *temporary* strategies The decreasing ORR phase with H higher than 15-25min does not affect the four strategies similarly. The *temporary* ones are the less impacted, with stabilization of ORR after a slight decrease. Providing more information to the dispatcher does not change the number of matches. The dispatcher has exploited the maximum amount of information possible. The additional notices included within the horizon are candidates for repositioning in the *temporary* strategies. However, only idle AVs (with no service mission) are candidates for temporary matches. Since the fleet is not oversized, most AVs are occupied with a service mission at each instant. No supplementary idle AV is available for being matched with the additional prior notices. The value of H from which stabilization happens depends mainly on fleet size (oversupply allows to keep more AVs available for repositioning missions).

For *permanent* strategies, we do not observe the same ORR stabilization when H increases over H=25min. Between H=25min and H=60min, *permanent-impatient* looses 254 matches and *permanent-productivist* looses 307 matches.

With *permanent-productivist*, extending H too much favours AMoD and puts customers at a disadvantage We find that for $H \geq$ 20min, TDTS and TDTE decrease. AVs provide less repositioning effort and serve fewer kilometers. In parallel, G reaches its minimum at H=20min and then increases. AMoD is more focused on central zones and serves shorter and closer requests. Extending the horizon makes more of these requests visible to the AMoD, and AVs usually sent in U_3 are kept in U_1 instead. As more and more AVs desert U_3 to wait for new customers in U_1 and U_2, the number of matches decreases. The average number of waiting AVs triples between H=20min and H=60min.

Meanwhile, P remains between 10950 and 11500 EUR, higher than other strategies. The maximum profit reached is 11471 EUR for H=25min. By focusing more on shorter and closer requests while ignoring more distant and longer rides, AMoD profit increases until the time lost due to AVs arriving ahead of time becomes disadvantageous.

Extending H too much is negative for ORR but dummy for G and P with *permanent-impatient* Both TDTS and TDTE grow at the same pace. AVs serve fewer customers but make more repositioning efforts for longer rides. The time to achieve a mission, from repo-

sitioning to drop-off, is greater, and AVs are monopolized longer for a given mission. Moreover, the number of candidates for a match increases with a wider horizon. Match being permanent, more AVs are necessary to answer a batch of requests. The average number of idle AVs is divided by two between H=20min and H=60min. With an over-extended horizon, *permanent-impatient* AMoD cannot maintain ORR. It assigns AVs with distant requests characterized by a far earliest pick-up time and requiring enough repositioning effort to allow the just-in-time arrival of AV at the pick-up point. These far requests, associated with longer rides, are profitable enough to be prioritized in the optimal matching resolution.

A wider horizon does not negatively impact P, nor G. P variation range is relatively narrow (10070-10289 EUR) since TDTS and TDTE increase jointly: repositioning efforts and service distance benefits are balanced. G maintains around 0.055 from H=30min.

The *temporary-impatient* strategy is the worse for P due to unproductive repositioning For *temporary-impatient*, we notice that TDTE keeps increasing while TDTS stabilizes from H=15min. AVs are making more repositioning efforts without any reward in terms of TDTS. Moreover, counting the average number of idle AVs shows that more AVs are running (serving or repositioning) during the simulation. Consequently, the profit curve is the worst among all strategies, reaching a minimum of 8489 EUR. A temporary match with a distant request requires more than αH=10min to be achieved. Then, it is put back into play and may undergo interruption. For $H \geq$ 15min, 80% of repositioning missions are interrupted. The repositioning AV is either permanently matched with another request than its previous target or assigned another repositioning target. The *temporary* strategy is at the origin of unproductive repositioning and sub-optimality in dispatching decisions here.

Repositioning in *temporary-productivist* is more flourishing and leads to a satisfying and stable P for the supplier but higher spatial disparities in ORR For *temporary-productivist*, TDTS, TDTE, and P are similar to *permanent-productivist* until H=10min. From then, the three indicators and G stabilize, letting this strategy below *permanent-productivist* in terms of P. The maximal profit reached is the second best (10825 EUR). We note that using the *productivist* utility rather than the *impatient* one allows reaching a higher ratio of repositioning missions achieved (80-90%). Indeed, the repositioning missions are shorter, so they have fewer occasions to be interrupted.

The *permanent* strategies are better for improving TWT and mean pick-up time as H increases TWT decreases as H increases. The *permanent* strategies allow reaching smaller TWT than *temporary*. They lead to a great reduction of the mean pick-up time, which starts around 12min for $H = 0$ and is lower than 2min at $H = 60$ for both *permanent* strategies. If the number of matches decreases, the quality of service for matched travelers largely increases.

The *temporary* strategies also have similar TWT profiles: it starts by decreasing, then stabilizes just as other indicators. The mean pick-up time reaches 8min (resp. 10min) for $H = 15$ and remains around 8min30s (resp. 12min) for a wider horizon under the *temporary-productivist* (resp. *temporary-impatient*) strategy.

The advantage of *productivist* utility definition for the supplier is evident in our environment: waiting passively in the center is a more profitable strategy for AMoD compared to actively looking for more distant rides. This behavior causes higher disparities in terms of ORR (G).

4.4.2.4 Extension to traveler-specific tolerance to wait

In this section, we wonder how strategies adapt to more impatient travelers. We go further in the analysis of travelers-side indicators. The *temporary-impatient* strategy is excluded, given its inability to maximize profit due to unproductive repositioning movements.

In this section, the maximum waiting time for being picked up is not constant but traveler-specific (w_r). We choose to approximate the tolerance to wait of each traveler based on available alternatives on the transportation network. Equation 4.9 corresponds to a simple behavior: when the traveler's waiting cost reaches the extra cost of not using AV during her trip, she cancels her request and diverts to an alternative itinerary. We assume that after having canceled once, a traveler will not try to use AV later in her journey: she follows the alternative path until her final destination.

$$w_r = \frac{c_r^{\bar{\mathrm{a}}} - c_r^{\mathrm{a}}}{\beta_{\mathrm{wait}}} \tag{4.9}$$

where c_r^{a} is the travel cost for the traveler associated with r on her currently followed itinerary (starts from the arrival node of a *pick-up* edge), $c_r^{\bar{\mathrm{a}}}$ is the travel cost for the traveler associated with r on the best itinerary between her current location and her final destination that does not rely on AMoD.

Figure 4.11 represents the mean maximum waiting time per service zone. It shows smaller values in zones better supplied by transit. Travelers who emit a request from a corner zone of U_3 are more tolerant to wait (16min) than travelers who emit a request from the central zone (7min). The more a traveler is tolerant to wait, the more she depends on AVs. Here, we investigate the inequalities in terms of quality of service. We wonder to what extent inequalities in quality of service are related to AV dependency.

Figure 4.12b shows that mean pick-up time is always lower in the traveler-specific maximum waiting time scenario (w_r). It is consistent since travelers are more demanding ($w_r \leq W$). The difference in mean pick-up time with W scenario reduces as H increases for *permanent* strategies while it stays constant for the *temporary-productivist* strategy. Figure 4.12a shows that with a short horizon, AMoD does not succeed in answering as many requests as in the W scenario. When horizon length is sufficient (from 15-25min), P for w_r and W scenarios are similar. It highlights the resilience of the approach under tougher timing constraints.

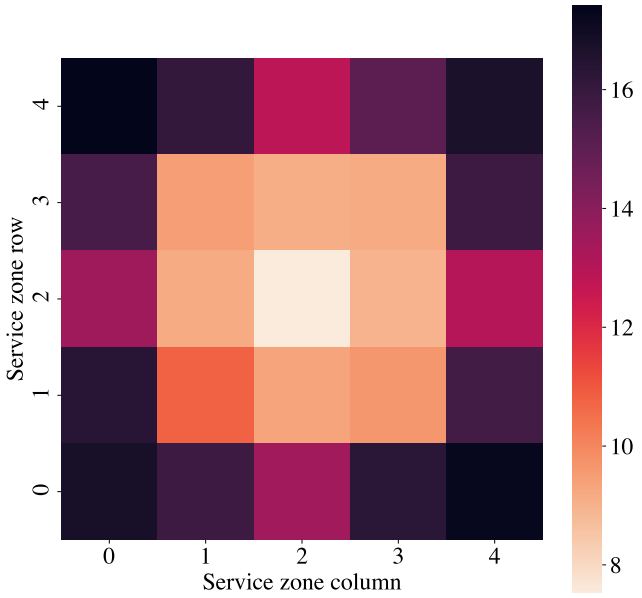
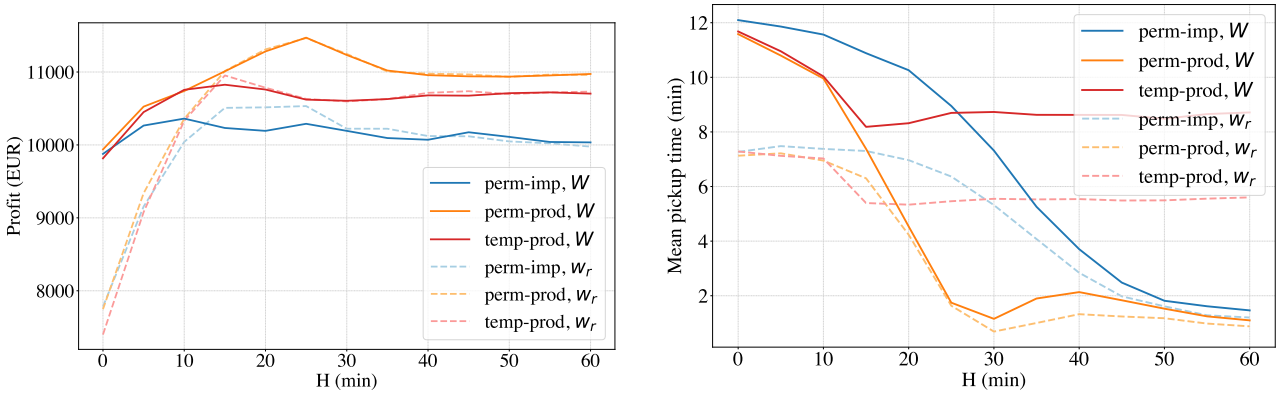


FIGURE 4.11 – *Mean maximum waiting time (w_r) in minutes per service zone.*



(a) *P* (b) *Mean pick-up time*

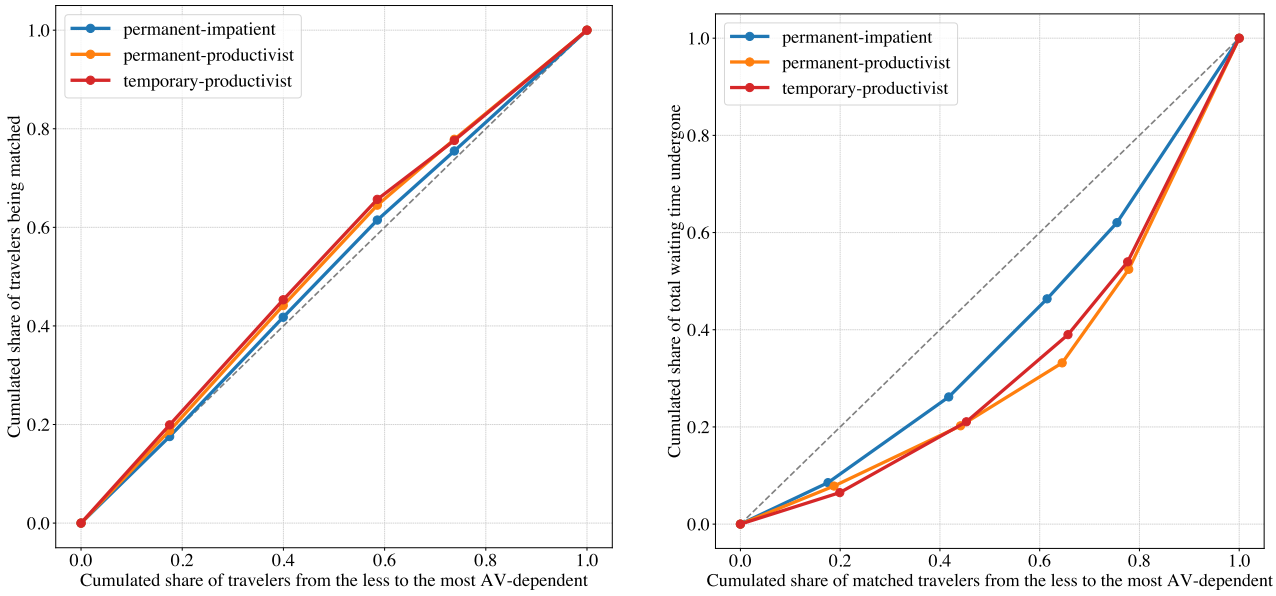
FIGURE 4.12 – *Sensitivity of H with homogeneous (W) and traveler-specific (w_r) maximum waiting time for being picked up.*

Figure 4.13 presents the concentration curves Wagstaff *et al.* (1991) for AMoD quality of service. To obtain it, we have defined five classes of AV-dependency from the less dependent to the most:

1. $w_r \leq 7$
2. $8 \leq w_r \leq 10$
3. $11 \leq w_r \leq 13$
4. $14 \leq w_r \leq 16$
5. $17 \leq w_r$

On the x-axis, it plots the cumulative proportion of travelers (all on 4.13a, only those that have been matched on 4.13b), beginning with the less AV-dependent, ending with the most AV-dependent. On the y-axis of Figure 4.13a (resp. Figure 4.13b), it plots the cumulative share of travelers being matched (resp. of total waiting time for pick-up). If matches (resp. waiting times) are equally undergone across classes, the concentration curve coincides with the diagonal. The degree of inequality can be represented by the concentration index, which equals 0 for perfect equality. It is worth -1 (resp. 1) for a perfect inequality to the advantage of the less AV-dependent, i.e., all matches are attributed to the less AV-dependent (resp. the most AV-dependent travelers undergo all the waiting time imposed by AMoD).

Table 4.3 presents the concentration indices for the three strategies with H=25min. On the concentration curves, one can see that all strategies favor the less Av-dependent travelers in terms of matches and pick-up time. The *productivist* utility leads to more inequalities. For the *permanenet-productivist* strategy, for example, the 50% of travelers the less AV-dependent attract nearly 60% of the matches. The 20% of matched travelers the most AV-dependent undergo almost half of the total realized pick-up time. Meanwhile, the 20% of matched travelers the less AV-dependent undergo only 8% of this quantity.



(a) *Concentration curve for the matches*

(b) *Concentration curve for the pick-up time*

FIGURE 4.13 – *Equity of the operational strategy in terms of quality of service (matches and pick-up time) for each strategy with H=25min.*

Strategy	**Concentration index**	
	Matches	**Pick-up time**
permanent-impatient	-0.025	0.209
permanent-productivist	-0.059	0.353
temporary-productivist	-0.072	0.339

TABLE 4.3 – *Concentration indices per dispatching strategy with H=25min.*

4.4.2.5 Influence of fleet size

Section 4.4.2.3 has shown that a horizon length between 15min and 25min improves AMoD-side and travelers-side benefits. However, we have stuck to a slightly undersized fleet. In this section, we question the ability of our batch-matching over horizon approach to manage the fleet correctly under different fleet sizes in comparison with an event-based matching policy with simple rules. The latter policy provides lower bounds for P and ORR indicators. It informs how an undersized, well-sized, and oversized fleet can be optimally managed without knowledge of future demand.

The event-based policy implemented is a *nearest-vehicle/nearest-open-request* heuristic. The set of known requests at t (R) only comprises requests r satisfying $e_r \leq t$. Each time a new request r is received, the set V_r is built. V_r contains all available vehicles candidate to be matched with r, i.e., able to arrive at r pick-up point on time, as defined by equation 4.10. If this set is not empty, v^* is matched with r (equation 4.11). If empty, r remains unmatched and is added to R_O.

$$V_r = \{v \in V_O \mid L_{r_k|P_{R_v}} + t_{r_k} + t_{r_k,r} \leq l_r \text{ if } |R_v| > 0,\ t + t_{r_0,r} \leq l_r \text{ otherwise}\} \tag{4.10}$$

$$v^* = \underset{v \in V_r}{\arg\min}(d_{r_k,r}) \quad \text{where } k = 0 \text{ if } |R_v| = 0 \tag{4.11}$$

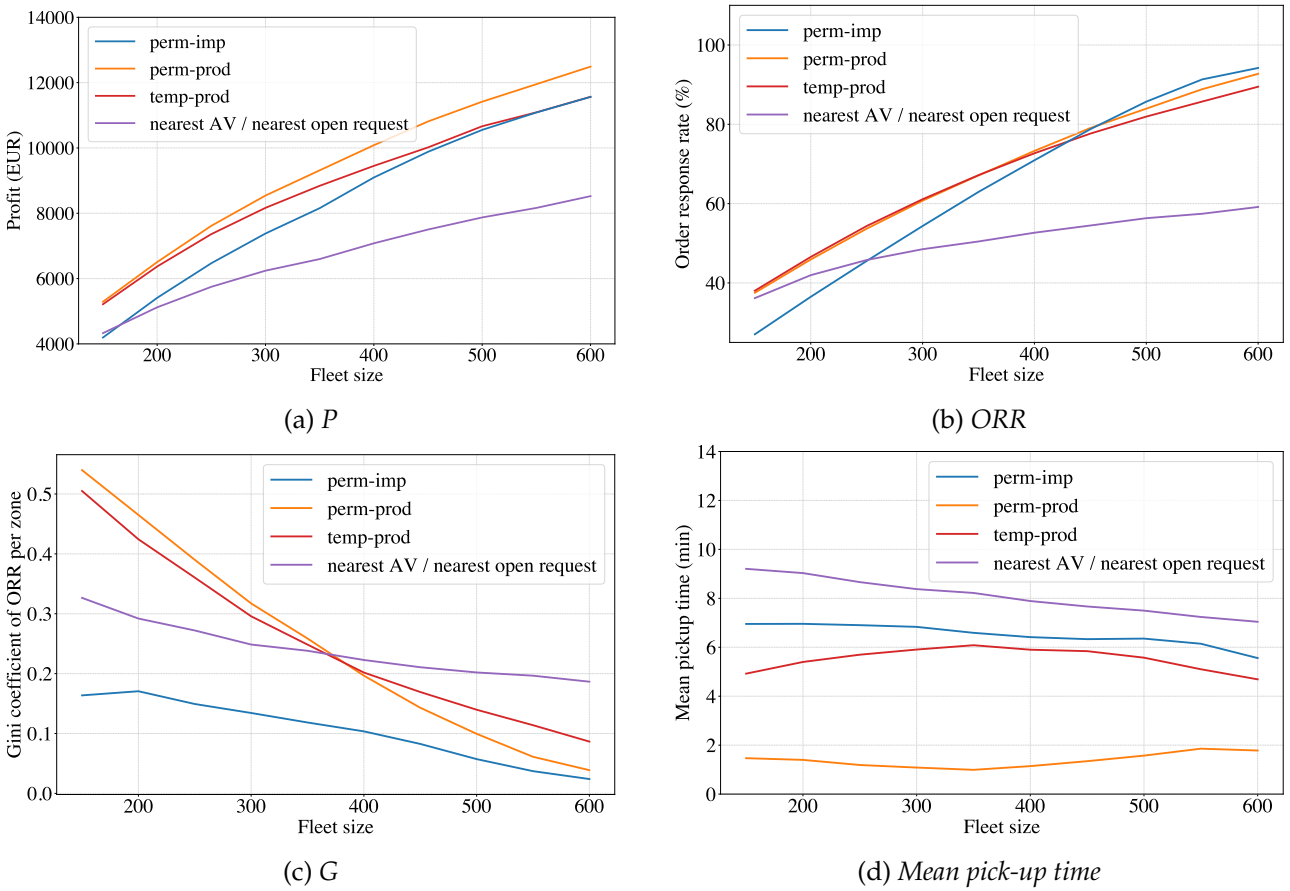
Each time a vehicle v drops off a customer, the set R'_v is built following equation 4.12. R'_v contains all open requests candidate to be matched with v, i.e., with a latest pick-up time compatible with v's service plan. If it is not empty, r^* is matched with v (equation 4.13). If it is empty, v is not matched.

$$R'_v = \{r \in R_O \mid L_{r_k|P_{R_v}} + t_{r_k} + t_{r_k,r} \leq l_r \text{ if } |R_v| > 0,\ t + t_{r_0,r} \leq l_r \text{ otherwise}\} \tag{4.12}$$

$$r^* = \underset{r \in R'_v}{\arg\min}(d_{r_k,r}) \quad \text{where } k = 0 \text{ if } |R_v| = 0 \tag{4.13}$$

Figure 4.14 presents the results for a traveler-specific maximum waiting time. When the fleet is oversized, many idle AVs fail to answer the remaining unmatched requests. In our instance, the lack of horizon prevents idle vehicles, mostly located in the central zone, from satisfying the timing constraints of corner zones' requests. The horizon allows the other three strategies to take advantage of supplementary resources: ORR keeps increasing at a high rate till the maximum fleet size. On average, for 600 AVs, there are 380 idle AVs with the event-based heuristic, 90 with the *permanent-impatient*, 60 with the *permanent-productivist*, and 160 with the *temporary-productivist*. The ORR (resp. P) gap is around 34% (resp. 3965 EUR). Note also that as fleet size increases, the gap between *permanent-impatient* and *permanent-productivist* strategies narrows.

When the fleet is undersized, the event-based heuristic gets closer to the other three strategies in terms of P and ORR. It even overcome the *permanent-impatient* strategy below 250 AVs.



(a) *P* (b) *ORR*

(c) *G* (d) *Mean pick-up time*

FIGURE 4.14 – *Sensitivity of fleet size on P, ORR, G, and mean pick-up time with traveler-specific maximum waiting time and H=25min under batch-matching and event-based matching policies.*

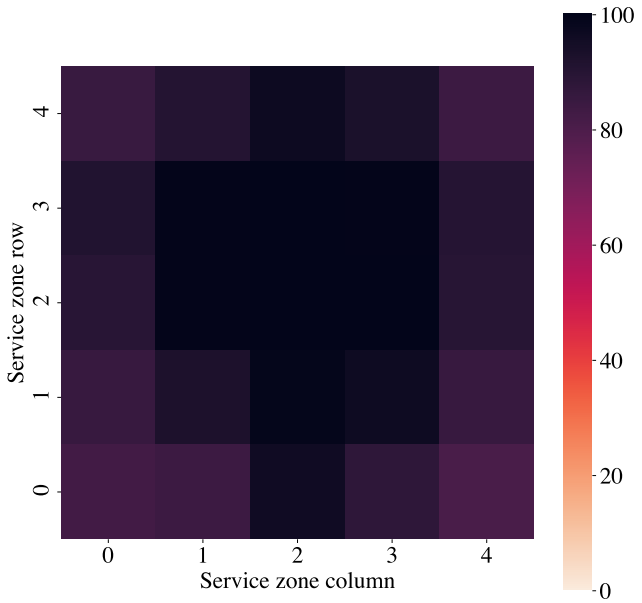


FIGURE 4.15 – *ORR per zone for permanent-productivist strategy with 600 AVs (H=25min).*

Regarding G, the event-based heuristic remains between 0.18 and 0.33. It is better than

the *productivist* strategies for a fleet containing less than 350 AVs but worse for greater fleets. The tendency to focus on less distant and shorter rides highlighted for *permanent-productivist* is close to what the event-based heuristic emulates. For batch-matching over horizon approach, when more AVs are available, ORRs of corner zones increase, and G is almost null, as shown in figure 4.15.

For the event-based heuristic, the mean pick-up time is always higher than for the other three strategies. It decreases from 9.2min for 150 AVs to 7min for 600 Avs, but this is at the price of numerous idle AVs. The *permanent-productivist* strategy is the best regarding mean pick-up time with a variation range extending from 1min to 2min.

4.4.2.6 Sensitivity to a naive pricing scheme

As highlighted above, *permanent-productivist* is the strategy maximizing the profit of AMoD, but it does to the detriment of equity among travelers under a limited fleet size. This section tests the sensitivity of two naive pricing schemes on the system behavior. These pricing schemes aim at restoring equity under the *permanent-productivist* strategy by taxing or subsidizing pick-ups per service zone. We test two intuitive pricing schemes:

- **PSC1**: The regulator provides subsidies to AVs for picking up a traveler in an AV-dependent zone and taxes AVs for picking up a traveler in a non-AV-dependent zone. Subsidies and prices are proportional to (i) the AV-dependency of the zone, (ii) the distance of this zone from the center of the city. Figure 4.16 shows prices per zone.
- **PSC2**: Subsidies are the same as in PSC1, but there is no tax.

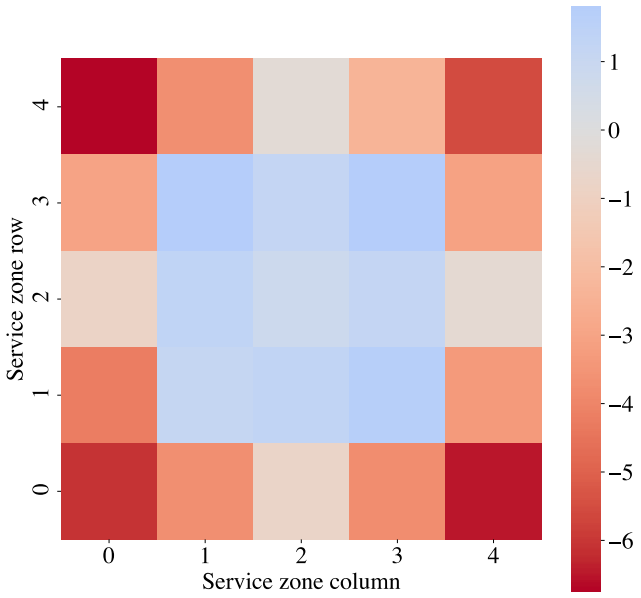
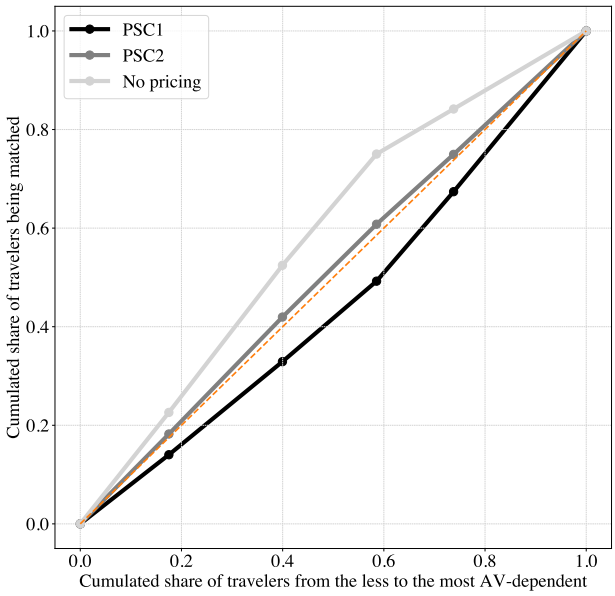


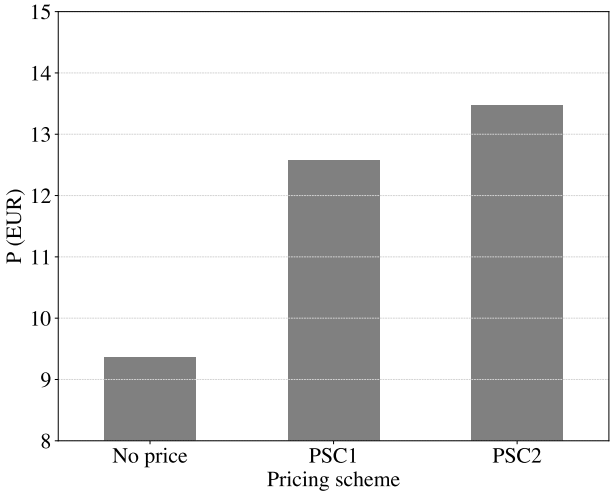
FIGURE 4.16 – *Pricing scheme 1 (PSC1) where taxes are positive and subsidies are negative (EUR).*

Figure 4.17 shows that compared to no pick-up price scenario, PSC2 almost reaches equality in terms of matches while PSC1 brings equity. In the latter scenario, the 40% travelers more AV-dependent get 50% of the matches. Due to the definition of schemes providing

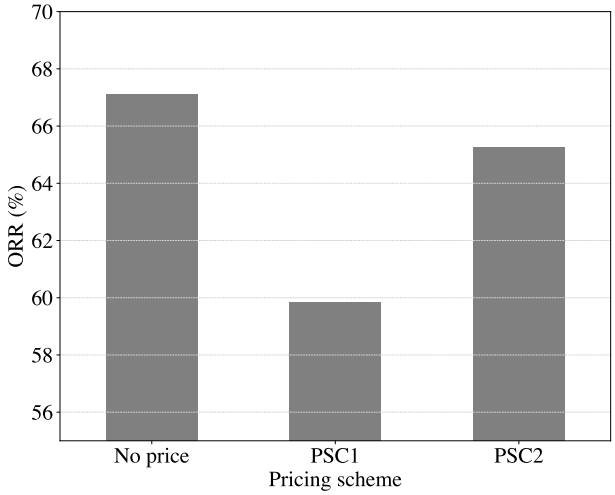
more subsidies than collecting prices, P increases by 3221 EUR with PSC1 and 4124 EUR with PSC2 while ORR slightly decreases.



(a) *Concentration curves for the matches under no pricing, PSC1 and PSC2.*



(b) *P*



(c) *ORR*

FIGURE 4.17 – *Effect of the naive pricing schemes on equity in matches, P and ORR for permanent-productivist strategy, traveler-specific maximum waiting times, 350 AVs and H=25min.*

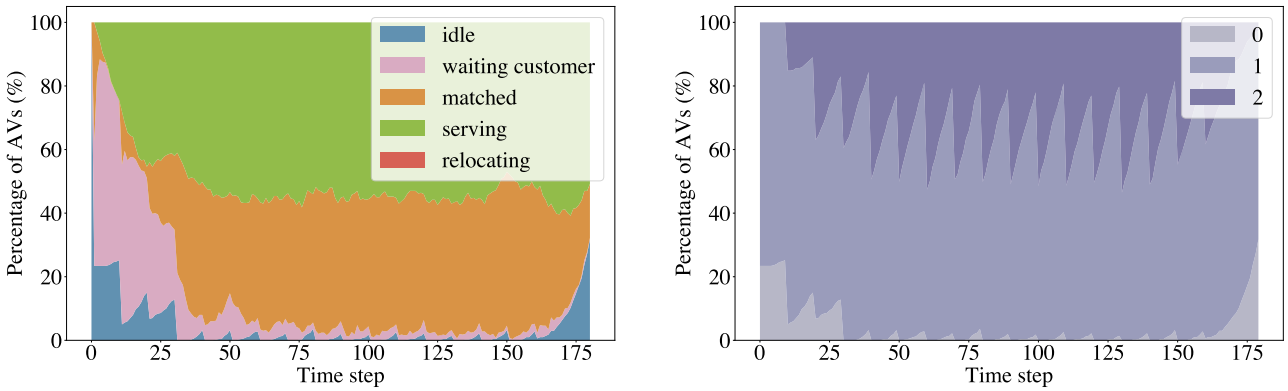
4.5 Discussion

4.5.1 One by one routing limitation

The biggest limitation of our batch-matching over horizon approach lies in constraint 4.8d of the optimal matching problem. Each vehicle can be matched at most to one request per matching round. Routing is done sequentially by adding requests one by one to a vehicle's

plan.

This constraint may result in lost resources between two calls of the matching problem. As discussed in section 4.4.2.2, reducing the time step between two calls of the matching problem (αH) does not have much impact on ORR when all travelers have the same maximum waiting time. It is also the case for traveler-specific maximum waiting time. Figure 4.18a shows that the percentage of idle AVs remains low between two matching rounds during the stationary regime (between t=50min and t=150min). One reason for this is that AVs can have more than one mission in their service plan, so they pass into *matched* state after a drop-off and not into *idle* state (Figure 4.18b). Another reason is that 77% of the requests have a service time t_r higher than 10min (Figure 4.19) which is the chosen value for αH.



(a) *Percentage of AVs in each state over time.*

(b) *Percentage of AVs with 0, 1, and 2 missions in their service plan over time.*

FIGURE 4.18 – *Activity of AVs over time for the permanent-productivist strategy with 500 AVs, traveler-specific maximum waiting time and H=25min.*

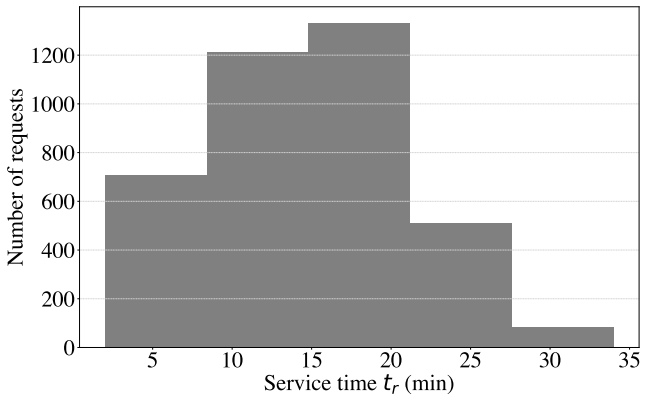


FIGURE 4.19 – *Distribution of service times (t_r).*

The constraint may lead to sub-optimal routes. Figure 4.20 presents a theoretical example for this. Two requests that could be added to one AV plan at the same matching round are respectively matched with one AV. They monopolize two AVs instead of one. A possible extension to the batch-matching policy proposed is to pre-build routes and then match

these routes with AVs, similarly to Alonso-Mora *et al.* (2017a). Appendix B.4 describes how the methodology proposed in the latter reference can be applied to our case with a profit-oriented operator and no ride-sharing.

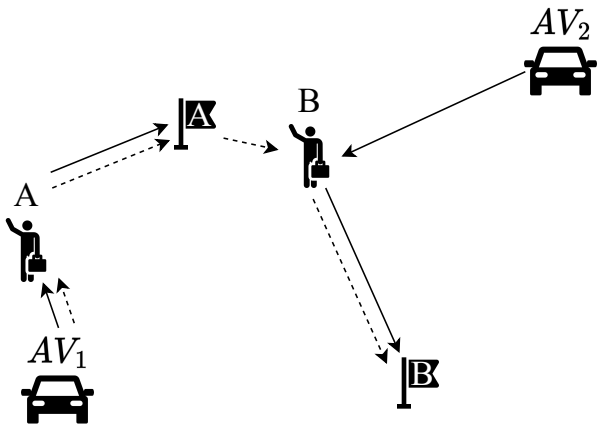


FIGURE 4.20 – *Example of potential sub-optimal routes due to the "at most one match per AV per matching round" constraint. The solid arrows represent the routes decided by the batch-matching over horizon policy when constraint 4.8d is enforced. The dashed arrows represent the optimal route when one AV can be matched with more than one request per matching round.*

4.5.2 Scalability

In the batch-matching over horizon approach proposed, two operations may be computationally challenging in large instances: the utility matrix computation and the resolution of the Integer Linear Program (ILP) defined in section 4.3.2.4.

The computation of the utility matrix has a $O(|V_O|\ |R_O|\ K)$ complexity for the insertion of a new request at the best index of the plan, and a $O(|V_O|\ |R_O|)$ complexity for the systematic insertion of a new request at the end of the plan. On top of choosing a low K value or a simplified insertion heuristic, several tricks can help reduce the computation time required to build the utility matrix. For each AV $v \in V_O$, one can limit the number of requests for which $u_{v,r}$ will be computed. First, one could only consider the requests with a pick-up point located in the vicinity of (i) drop-off points within P_{R_v} for the insertion at best index, (ii) last drop-off in P_{R_v} for the insertion at the end of the plan. It can be applied by defining an upper bound for the number of requests considered as candidates for a match (sort requests and select the first x) or a radius. The radius can be fixed or request-specific depending on p_r. If the (Manhattan) distance between the request pick-up point and the closest drop-off point in P_{R_v} leads to a higher repositioning cost than p_r, then do not consider r as a candidate for a match. Second, computations of utilities can be done in parallel. Third, the shortest paths between drop-off and pick-up points can be pre-computed.

ILP are NP-hard problems that state-of-the-art solvers can solve with the branch-and-cut method. Since we consider a unique profit maximization objective for AMoD, we can reduce the number of decision variables for the problem by setting $x_{v,r}$=0 for each (v, r) so

that $u_{v,r} < 0$. Here, we used CPLEX 20.1.0 with the default termination condition [2]. One can set a time limit instead. Using a 'good' solution as the root node of the branch-and-cut tree can also help reduce the computation time. This starting point solution can be computed with a naive matching policy (e.g., decide matches in decreasing utility order).

4.6 Conclusion

In this chapter, we have justified the choice for a batch-matching over horizon approach to model the profit-oriented operational policy of an AMoD operator. This choice relates to the context of pricing scheme design in a large urban area. We have described the dispatch algorithm in detail and proposed a simulation platform to test and analyze it under similar conditions as in chapter 5. Moreover, using a deterministic mode-route choice model on a multimodal network provided a realistic demand pattern and time window constraints. The analysis focused on the impact of the horizon length. We have shown that horizon is necessary to manage the fleet correctly and be sensitive to the pricing scheme implemented by the regulator. We highlighted the range of horizon length on which the batch-matching approach is still relevant. A 15-25min horizon length allows the AMoD operator to maximize its profit or ORR depending on its definition of utility. The approach performed well with patient and impatient travelers. Batch-matching performs better than a classic event-based policy, and the larger the fleet, the wider the gap. The system reacted as expected to a static naive pricing scheme designed to bring equity in matches across the urban area. This chapter's elements serve as a basis for chapter 5.

[2] MIP gap $\leq 0.01\%$ where MIP gap measures the progress toward finding optimality

5. THE PRICING SCHEME DESIGN PROBLEM: A SIMULATION-MULTI-OBJECTIVE OPTIMIZATION APPROACH

Some elements of this chapter have been presented in:

- Cortina, M., Khalesian, M., Leclercq, L. (2024). Multi-modal Traffic Management Optimization Using Gaussian Process and Pareto-based Multi-Objective Evolutionary Approach, In *Transportation Research Board 103rd Annual Meeting*, Washington DC, USA.

5.1 Motivation and objectives

In chapter 4, we have introduced and analyzed a profit-oriented operational strategy for AMoD. It has been tested with a simple simulation platform that ignores traffic dynamics and lacks flexibility in defining new mobility services with specific behaviors. To overcome these limitations, we switch now to a more powerful simulation platform called Multimodal network Modeling and Simulation [1](MnMS).

Developed by the LICIT-ECO7 laboratory, this open-source simulation platform matches the require modeling scale to study and design regulation policies for new mobility services. However, as any simulation tool, it is a black box that requires significant time to run. An efficient optimization technique is required to achieve the objective of part III, i.e., designing a pricing scheme that maximizes the benefits of intermodal AMoD.

Bayesian Optimization (BO) is a state-of-the-art technique for black-box optimization, i.e., problems for which (i) the objective function cannot be analytically defined, (ii) there is no information available about the mathematical properties of the objective function and the potential constraints, (iii) derivatives are not easily calculable, and (iv) evaluation is time-consuming (from several minutes to several hours). Widely used in engineering design (Priem, 2020) and machine learning (Wu *et al.*, 2019), BO has shown its efficiency on problems with up to 20 decision variables and noisy objective functions. It has been applied to several transportation problems, including transportation model calibration (Sha *et al.*, 2020), traffic signal control (Tay & Osorio, 2022), AV market share optimization (Fakhrmoosavi *et al.*, 2022), and congestion pricing (Huo *et al.*, 2023, Liu *et al.*, 2021).

Two studies using BO for (A)MoD design and regulation purposes are relevant to our work. Liu *et al.* (2019) develop a simulation-optimization framework to design, optimize,

[1]Source code is available at https://github.com/licit-lab/MnMS

and analyze MoD operations within a multimodal transportation system. In the simulation loop, they simulate transit and three MoD services with varying passenger capacities managed by a single operator and calculate a user equilibrium iteratively. In the optimization loop, they use BO to maximize the profit of the MoD operator with each MoD service fleet size and fare rules as decision variables. Dandl *et al.* (2021) develop a tri-level simulation-optimization framework to study the regulation of AMoD. They simulate transit and AMoD operations at the lower level and calculate a user equilibrium. At the second level, they use BO to optimize AMoD fleet size, distance-based fare, and fare scale factor to maximize the AMoD operator's profit. At the upper level, they use BO again to optimize regulation measures (parking fees, road toll per driven km, transit frequencies scale factor, and the number of AV licenses) to maximize social welfare, including travelers' utilities, transit, and AMoD profits, revenues from parking and tolls, and CO2 emissions cost.

Both studies present differences from our work. While Dandl *et al.* (2021) model congestion at the same scale as MnMS does, Liu *et al.* (2019) ignore traffic dynamics. These works focus on a few transportation modes: a set of MoD services (ride-hailing, ridepooling, microtransit) and subway as the only transit option for Liu *et al.* (2019), personal car, transit and AMoD for Dandl *et al.* (2021). They both ignore intermodal options and concentrate on the benefits of sharing rides. In contrast, we focus on the benefits of splitting door-to-door rides into intermodal ones. Moreover, both studies use BO to solve single-objective problems. Liu *et al.* (2019) tackle MoD operator profit maximization. Dandl *et al.* (2021) define a social welfare function as a weighted sum of travelers' utilities, transit, AMoD, regulator profits, and CO2 emissions cost. In practice, policymakers do not know the weights of such functions. They must evaluate several regulations under different priorities and find the best trade-off depending on contextual elements that cannot be captured.

For this reason, we adopt a multi-objective optimization approach in this chapter. We study Pareto fronts for several sets of objective functions and regulation prices. The Pareto fronts profiles deliver valuable information about the system and the best way to regulate it. While BO originally deals with only one well-defined objective function, the literature contains several adaptations of BO for the multi-objective context. Among them, the MOBOpt algorithm (Galuzio *et al.*, 2020) is a state-of-the-art Multi-Objective Bayesian Optimization (MOBO) algorithm that builds a surrogate model for each objective function and exploits them to sample points where launching a simulation is valuable. The sampling method tries to maximize the diversity of the non-dominated points found to produce the highest quality Pareto front possible. The application of MOBOpt to our problem allows us to derive several optimal regulation policies for maximizing the benefits of intermodal AMoD with regard to several sets of objectives.

This chapter has three purposes. First, it presents the agent-based simulation approach and describes how the AMoD operational policy studied in chapter 4 integrates with the MnMS platform. Second, it formulates the pricing scheme design as a multi-objective optimization problem and presents the solving method. Third, it analyzes the results for several sets of prices, several sets of objective functions, and discusses the performance of intermodal AMoD under the optimized regulations.

5.2 Notations for this chapter

TABLE 5.1 – *Notations for chapter 5.*

Notation	Definition
	Multi-Objective Optimization Problem
$\mathscr{P}$	Regulation policy (a certain set of regulation prices)
$x^{\mathscr{P}}=x$	Decision variable, vector of prices values
Ω	Search space ($\Omega \subset \mathbb{R}^{\lvert\mathscr{P}\rvert}$)
f_i	i^{th} objective function of the problem
n	Number of objectives in the problem
	Multi-Objective Bayesian Optimization
K_{init}	Number of initial points
$\mathcal{D}_{f_i}^{(q)}$	Set of the first q observed/evaluated points
$x^{(q)}$	q^{th} point observed
$\Phi^{(q)}$, $\Psi^{(q)}$	Observed Pareto Front and Set after q evaluations
$\mathcal{GP}_i(\mu, k\lvert\mathcal{D}_{f_i}^{(q)}, x)$	Gaussian process fitted on the first q evaluations characterized by mean μ and kernel function k
$\hat{\mu}_{f_i}^{(q)}(x)$, $\hat{\sigma}_{f_i}^{(q)}(x)$	Estimation of f_i value at x by the Gaussian process, estimation of the associated prediction error
$K^{(q)}(x)^{\mathrm{T}}$	Covariance vector obtained by applying the kernel function to the observations and x
$K^{(q)}$	Covariance matrix of observations
$F_i^{(q)}$	Vector of the objective function i values at the observations
$\mu^{(q)}$	Mean vector at the observations
$\mathcal{C}_{5/2}$	Matérn 5/2 kernel function
I	Identity function
$\theta_0, ..., \theta_{\lvert\mathscr{P}\rvert}, \sigma$	Hyperparameters for the Matérn 5/2 and white kernels
x_j	j^{th} component of x
$\hat{\Phi}^{(q)}$, $\hat{\Psi}^{(q)}$	Estimated Pareto Front and Set after q evaluations
$x^{(q+1)}$, $\tilde{x}^{(q+1)}$	Next point to evaluate, mutated next point to evaluate
$\tilde{\Omega}$	Set of all possible mutated next point where one component of $x^{(q+1)}$ only can be mutated
$\delta_{\Omega}^{(q)}(x)$	Least distance between x and the q evaluated points
$\delta_{\Omega,\mu}^{(q)}$, $\delta_{\Omega,\sigma}^{(q)}$	Mean and standard deviation of all $\delta_{\Omega}^{(q)}(x)$, $x \in \hat{\Psi}^{(q)}$
$\delta_{f}^{(q)}(x)$	Least distance between $\left[\hat{\mu}_{f_i}^{(q)}(x)\right]_{i\in[\![1,n]\!]}$ and the q evaluations
$\delta_{f,\mu}^{(q)}$, $\delta_{f,\sigma}^{(q)}$	Mean and standard deviation of all $\delta_{f}^{(q)}(x)$, $x \in \hat{\Psi}^{(q)}$

Continue on the next page

TABLE 5.1 – *Notations for chapter 4 (continued).*

Notation	Definition
w	Weight of the search space over the objectives space for the next point selection
	Simulation
dt_{match}	Time step for AMoD matching
H	Horizon length for travelers departures anticipation and AMoD matching
dt_{aff}	Time step for the affectation of travelers on a mode-route
dt_{flow}	Time step for vehicles and users movement on the network
t	Current time
t_{start}, t_{end}	Start and end time of the studied period
	Mode and route choice
o_i, d_i	Origin and destination of traveler i
t_i^{dep}	Departure time of traveler i
M	Set of all modes considered in this chapter (**C** for car only, **A** for AMoD only, **P** for public transportation only, **AP** for AMoD and public transportation, **CP** for car and public transportation)
M_i	Set of modes available for traveler i
m, m_j	Index of mode, index of the jth mono-modal sub-mode composing mode m
π_i^m	Probability for traveler i to choose mode m
C_i^m	Generalized cost for i to travel on the optimal route of mode m
$t_{i,\text{wait}}^m$, $t_{i,\text{walk}}^m$	Total waiting/walking time on i's optimal route for mode m
t_{i,m_j}^m, d_{i,m_j}^m	In-m_j-vehicle time/distance on i's optimal route for mode m
η_i^m	Total number of transfers occurring on i's optimal route for mode m
η_{i,m_j}^m	Total number of transfers toward a m_j vehicle occurring on i's optimal route for mode m
β_i^{drive}	Value of time of the driving activity for traveler i, similar notations for the AV riding (rideA), public transportation riding (rideP), walking (walk), and waiting (wait) activities
$\gamma_i^{\text{transfer}}$	Transfer penalty for i to change vehicle
p_i^m	Total regulation price that applies to i on mode m
c_{d}^{C}, c_d^{A}	Operational and ownership cost of car, operation cost of AV
f_0^{A}, f_d^{A}, f_t^{A}	Base/Distance-based/Time-based fare of AMoD
f_0^{P}, f_d^{P}	Public transportation ticket price, additional distance-based train fare
	AMoD matching

Continue on the next page

TABLE 5.1 – *Notations for chapter 4 (continued).*

Notation	Definition
r_i=r	Request issued by traveler i to AMoD (both notations are equivalent in the chapter, note that i may have issued several requests to AMoD, the notation refers to any request issued by i)
e_r, l_r	Earliest/Latest pick-up time for request r
$t_{\text{wait}}^{\text{A,max}}$	Maximum cumulative waiting time for AMoD over the whole trip
R_{O}, $R_{\text{O,H}}$	Set of open requests, set of open requests within the horizon
V, V_{O}	Set of AVs composing the fleet, set of available AVs
R_v	Set of requests that have been matched with AV $v \in V$
P_{R_v}	Activity plan of AV $v \in V$
$P^*_{R_v \cup \{r\}}$	New activity plan of AV v that includes r pick-up and serving activities in the end
$u_{v,r}$	Utility of a match between v and r for AMoD operator
$t^{\text{end}}_{P_{R_v}}$	Estimated end time of plan P_{R_v}
K_{match}, K	Maximum number of iteration for the AMoD iterative batch-matching over horizon algorithm, maximum number of requests within an AV plan
	Roads
v_z	Mean speed of traffic within reservoir z
L_z, N_z,V_z	Total length of roads, maximum number of running vehicles, free-flow speed within zone z
N_z^{C1}, N_z^{C2}, V_z^{C1}, V_z^{C2}	Critical accumulations and speeds in z
	Multimodal network
v_{walk}, v_{train}, v_{metro}	Walking speed, commercial speed of train, and metro
h_{train}, h_{metro}, h_{bus}	Headway of train, metro, bus lines
N_{A}	AMoD fleet size
$\rho_{\text{acc/egr}}$	Maximum walking distance to access the network from origin and egress the network to destination
ρ_{transfer}	Maximum walking distance for transfer (mode or vehicle change)
	Demand
s	Scaling factor for users, cars, and AVs
θ_{C}	Percentage of travelers that have access to **C** mode
	Emissions
d_z^{C}, d_z^{A}	Total distance traveled by cars, Avs (serving and empty) within zone z

Continue on the next page

TABLE 5.1 – *Notations for chapter 4 (continued).*

Notation	Definition
e_C, e_A	Cars emissions factor function, AVs energy consumption factor function
E_{CO2}	Total CO2 emissions during the studied period
	Prices
p_{TA}	Price applied to travelers for a door-to-door AV ride
p_{TAP}	Price applied to travelers for an intermodal AV ride
p_{TC}	Price applied to travelers for a car trip
p_{AA}	Price applied to AVs for serving a door-to-door AV ride
p_{AAP}	Price applied to AVs for serving an intermodal AV ride

5.3 Problem formulation and solution approach

5.3.1 Problem formulation

To prepare for the deployment of a fleet of AVs operated by a private AMoD operator, the transportation authority of an urban area considers implementing a regulation policy based on picing. The policy aims to prevent competition between the new AMoD service and the existing public transportation, i.e., to steer the system to a certain collective optimum where AMoD and public transportation complete each other and are competitive against personal cars. The regulator has several considerations, such as lowering the environmental cost of the morning commute in the system, getting as close as possible to a revenue-neutral pricing scheme by re-investing the money earned through taxes as subsidies, and ensuring the fairness of the regulation.

Let $\mathscr{P}$ be a regulation policy, i.e., a certain set of regulation prices that apply to the system. Under $\mathscr{P}$, the decision variables of the regulator correspond to a vector $x^{\mathscr{P}}$ of prices values. To simplify the notations, we will note x in the rest of the chapter and specify the elements in $\mathscr{P}$ when necessary. The search space for x is noted $\Omega \subset \mathbb{R}^{|\mathscr{P}|}$. Regulator's considerations correspond to the n objective functions of a multi-objective optimization problem formulated by:

$$\min_{x \in \Omega} f_1(x), f_2(x), ..., f_n(x) \tag{5.1}$$

where $f_i : \Omega \mapsto \mathbb{R}, \forall i \in \{1, .., n\}$.

Solving such a problem implies finding the set of non-dominated points, also called the Pareto Set (PS). These solutions cannot be improved in any of the objectives without degrading at least one of the other objectives. Formally, $x \in \Omega$ dominates $x' \in \Omega$ if: (i) $\forall i \in \{1, ..., n\}$, $f_i(x) \leq f_i(x')$, and (ii) $\exists i \in \{1, ..., n\}$, $f_i(x) < f_i(x')$. The corresponding objective functions values are the Pareto Front (PF). The PF provides knowledge on how the objectives are conflicting. The regulator can exploit it to find the best trade-offs between its conflicting objectives. The PS guides the regulator to determine the proper values for each price of the policy.

In this chapter, we aim to derive the PF and PS for several relevant regulation policies and sets of objective functions.

5.3.2 Solution approach

The study uses a simulation-multi-objective optimization approach to derive the PF and PS, as represented in Figure 5.1. As stated in chapter 1, agent-based simulation allows detailed modeling of the interactions between travelers, the AMoD operator, and the regulator. Travelers and AMoD operator behaviors are simulated given the regulation policy implemented by the regulator. Each agent pursues its own objective.

The AMoD operator maximizes its profit per time unit by making dispatching decisions every dt_{match} in the light of the set of open requests within a horizon of length H, the set of available AVs, and the current traffic conditions. More details about the AMoD dispatching strategy are provided in section 5.4.2.2.

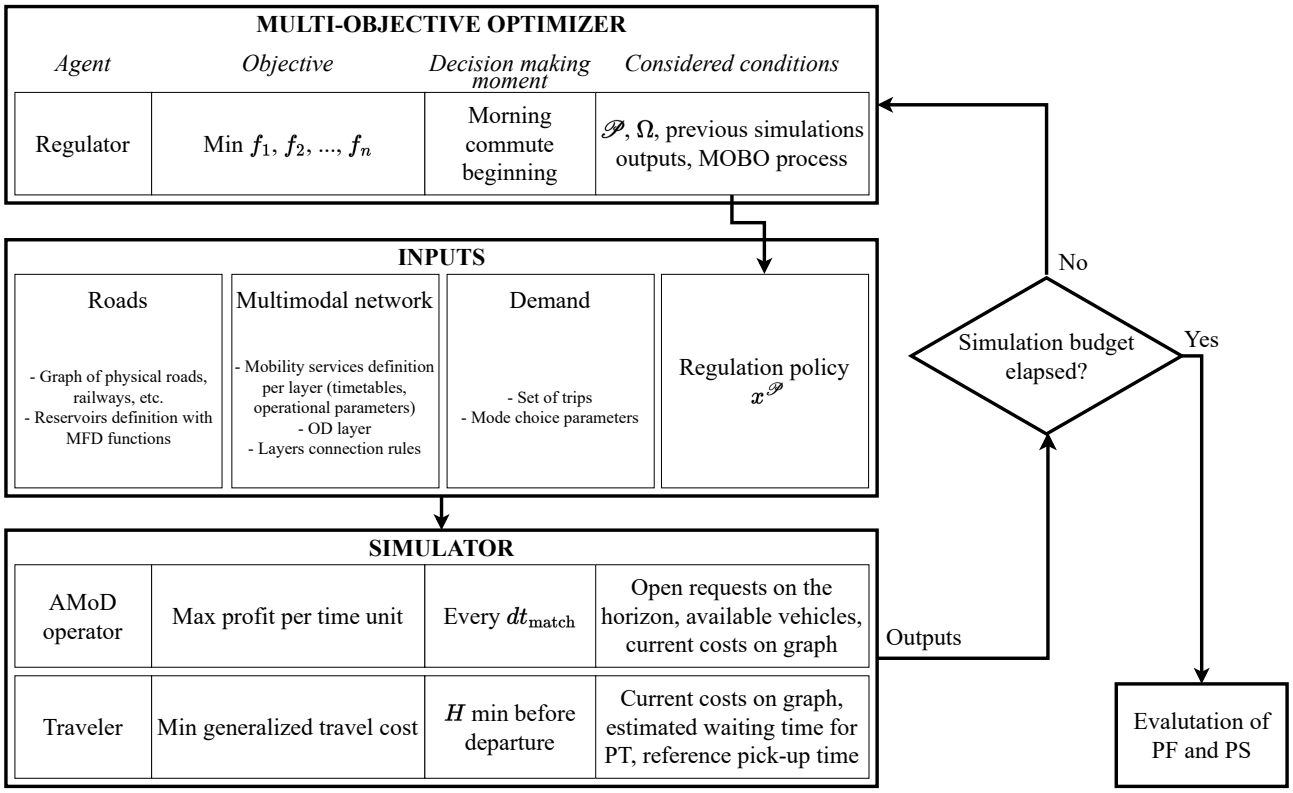


FIGURE 5.1 – *Agent-based simulation and optimization approach.*

Each traveler minimizes her generalized travel cost by choosing a certain mode and route H time units before her departure in light of the current traffic conditions and simulation parameters. More details about travelers' mode and route choices are given in section 5.4.1.

The regulator stands at the upper level of the approach. It aims to minimize several objectives jointly by choosing the regulation prices values at the beginning of the morning commute. Note that the regulation is static: prices are constant for the whole morning peak period. An efficient MOBO algorithm is used to choose the next x vector to evaluate through simulation. This iterative algorithm seeks to improve the observed PF deduced from all evaluated (simulated) vectors. The procedure stops once the simulation budget has elapsed. Once it has, the final PF and PS are carefully analyzed.

5.4 Simulation model

Figure 5.2 represents the global flowchart of the simulation. It is characterized by two distinct time steps: dt_{aff} and dt_{flow}. The finest is dt_{flow} and corresponds to the frequency at which users and vehicles of the simulation are moved. The largest, dt_{aff}, is a multiple of dt_{flow} and corresponds to the frequency at which costs of the graph links are updated and travelers make their mode and route choices. In the following subsections, we provide more details for each step of the flowchart.

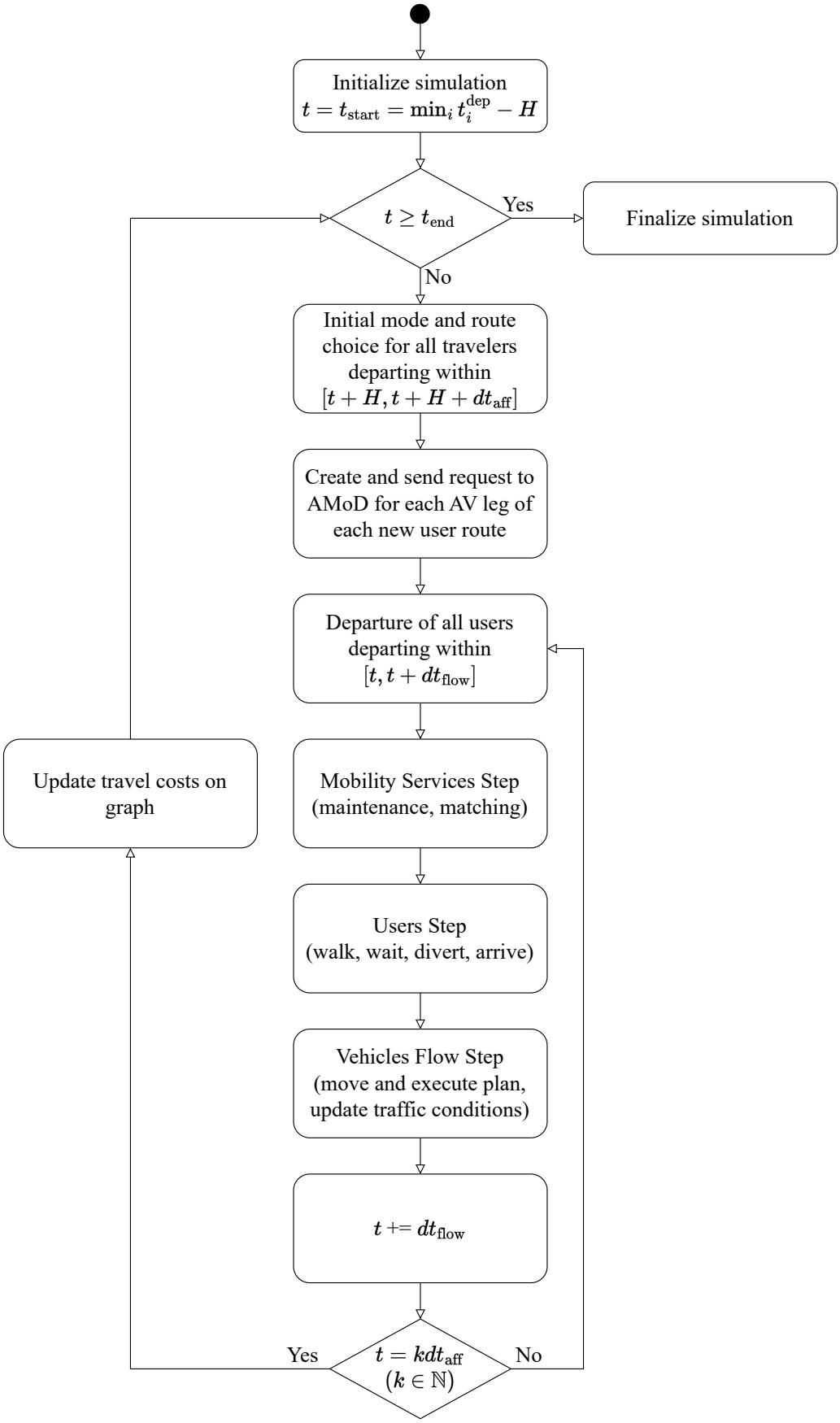


FIGURE 5.2 – *Simulation global flowchart.*

5.4.1 Mode and route choice

A traveler i looks to travel from an origin o_i to a destination d_i at the lowest generalized cost. For the sake of simplicity, we do not consider demand elasticity here, nor the impact of regulation measures and AV fleet behavior on departure times. In other words, i has a fixed departure time t_i^{dep} and definitely achieves its trips. To do so, she has access to a set of modes $M_i \subset M$. M designates the set of all modes considered in this study: personal car (**C**), AMoD only (**A**), public transportation and walk only (**P**), AMoD and public transportation (**AP**), car and public transportation (**CP**).

To select one mode in M_i, i uses a deterministic (5.2) model:

$$\begin{cases} \pi_i^{m^*} = 1 \\ m^* = \underset{m \in M_i}{\arg\min}\, C_i^m \end{cases} \tag{5.2}$$

where π_i^m is the probability for user i to choose mode m, C_i^m is the generalized cost for i to travel on the optimal route of mode m. Note that a more advanced model, such as a logit one, could have been considered.However, the deterministic mode choice model is chosen as a first step to validate the approach. In this way, one simulation setting leads to one value for each objective function. The application of MOBO in scenarios with non-deterministic choice models and hence, noisy objectives, is kept for future research.

By optimal path, we mean the path of minimum travel cost. To find the optimal route for each mode, the Dijkstra algorithm is used on the subgraph corresponding to this mode, with link-level travel costs as weight on the links. The quality of the routes found by the algorithm for intermodal modes (**AP** and **CP**) is verified before the mode selection to prevent unrealistic routes from being chosen. Two simple checks consist in (i) ignoring the intermodal mode m, using the two mono-modal modes m_1 and m_2, if the optimal route on m_1 or on m_2 uses walk only, (ii) ignoring the intermodal mode m if its optimal route passes several times by at least one node.

The generalized travel cost associated with a mode-route is defined in this study as the sum of the travel time cost, the mobility service monetary cost and the regulation monetary cost of this mode-route. Let us note:

- $t_{i,\text{wait}}^m$ the total waiting time on i's optimal route for mode m (including the potential waiting times for being picked-up by AV and waiting times at public transportation stops)
- $t_{i,\text{walk}}^m$ the total walking time on i's optimal route for mode m (including the access to the first road node of the route from the origin position, the egress from the last road node of the route to the destination position, and the potential transfers user travels by walking between two public transportation stops, between a road node and a public transportation stop or vice-versa)
- t_{i,m_j}^m and d_{i,m_j}^m the in-m_j-vehicle time and distance on i's optimal route for mode m where m_j is a mono-modal sub-mode of m

- η_i^m is the total number of transfers occurring on i's optimal route for mode m (where a transfer implies that user changes vehicle)
- η_{i,m_j}^m the number of transfers toward a m_j vehicle occurring on i's optimal route for mode m
- β_i^{drive}, β_i^{rideA}, β_i^{rideP}, β_i^{walk} and β_i^{wait} the user's values of time for the driving, AV riding, public transportation riding, walking, and waiting activities
- $\gamma_i^{\text{transfer}}$ the transfer penalty for i to change vehicle
- p_i^m the regulation price that applies to i for mode m

Travel costs are defined as follows:

$$\begin{aligned} C_i^{\text{C}} =& \beta_i^{\text{drive}} t_{i,\text{C}}^{\text{C}} + \beta_i^{\text{walk}} t_{i,\text{walk}}^{\text{C}} \\ &+ c_{\text{d}}^{\text{C}} d_{i,\text{C}}^{\text{C}} \\ &+ p_i^{\text{C}} \end{aligned} \tag{5.3}$$

where c_{d}^{C} is the full cost of car (operation and ownership) per distance unit.

$$\begin{aligned} C_i^{\text{A}} =& \beta_i^{\text{wait}} t_{i,\text{wait}}^{\text{A}} + \beta_i^{\text{rideA}} t_{i,\text{A}}^{\text{A}} + \beta_i^{\text{walk}} t_{i,\text{walk}}^{\text{A}} \\ &+ f_0^{\text{A}} + f_d^{\text{A}} d_{i,\text{A}}^{\text{A}} + f_t^{\text{A}} t_{i,\text{A}}^{\text{A}} \\ &+ p_i^{\text{A}} \end{aligned} \tag{5.4}$$

where f_0^{A}, f_d^{A} and f_t^{A} are the base, distance-based, and time-based parameters for AMoD fare scheme.

$$\begin{aligned} C_i^{\text{P}} =& \beta_i^{\text{wait}} t_{i,\text{wait}}^{\text{P}} + \beta_i^{\text{rideP}} t_{i,P}^{\text{P}} + \beta_i^{\text{walk}} t_{i,\text{walk}}^{\text{P}} + \gamma_i^{\text{transfer}} \eta_i^{\text{P}} \\ &+ f_0^{\text{P}} \end{aligned} \tag{5.5}$$

where f_0^{P} is the public transportation ticket price.

$$\begin{aligned} C_i^{\text{AP}} =& \beta_i^{\text{wait}} t_{i,\text{wait}}^{\text{AP}} + \beta_i^{\text{rideA}} t_{i,\text{A}}^{\text{AP}} + \beta_i^{\text{rideP}} t_{i,\text{P}}^{\text{AP}} + \beta_i^{\text{walk}} t_{i,\text{walk}}^{\text{AP}} + \gamma_i^{\text{transfer}} \eta_i^{\text{AP}} \\ &+ f_0^{\text{P}} \eta_{i,\text{P}}^{\text{AP}} + f_0^{\text{A}} \eta_{i,A}^{\text{AP}} + f_d^{\text{A}} d_{i,\text{A}}^{AP} + f_t^{\text{A}} t_{i,\text{A}}^{\text{AP}} \\ &+ p_i^{\text{AP}} \end{aligned} \tag{5.6}$$

where the user should pay the public transportation ticket every time she enters the public transportation network, i.e., direct transfers from one public transportation line to another can be achieved with the same ticket but indirect transfers (from one line to AV and from AV to another line) requires a new ticket.

$$\begin{aligned} C_i^{\mathrm{CP}} =& \beta_i^{\mathrm{wait}} t_{i,\mathrm{wait}}^{\mathrm{CP}} + \beta_i^{\mathrm{drive}} t_{i,\mathrm{C}}^{\mathrm{CP}} + \beta_i^{\mathrm{rideP}} t_{i,\mathrm{P}}^{\mathrm{CP}} + \gamma_i^{\mathrm{transfer}} \eta_i^{\mathrm{CP}} + \beta_i^{\mathrm{walk}} t_{i,\mathrm{walk}}^{\mathrm{walk}} \\ & + c_{\mathrm{d}}^{\mathrm{C}} d_{i,\mathrm{C}}^{\mathrm{CP}} + f_0^{\mathrm{P}} \\ & + p_i^{\mathrm{CP}} \end{aligned} \tag{5.7}$$

During the initial mode and route choice step, the in-vehicle times t_{i,m_j}^m of i on the optimal routes of modes $m \in M_i$ are estimated based on the current traffic conditions. At worst, a traveler decides based on the traffic conditions $H + dt_{\mathrm{aff}}$ minutes before her departure time. In the numerical application, we choose H and dt_{aff} values to keep this delay reasonable.

The wait times $t_{i,\mathrm{wait}}^m$ of i on the optimal routes of modes $m \in M_i$ are estimated based on the simulation parameters. The waiting time at a public transportation stop is estimated to be half the headway of the line serving this stop. The wait time to be picked up by an AV considered during the initial mode and route choice step is null for all users for the whole studied period. Indeed, in our model, travelers request AVs in advance with specific time windows for the pick-ups. The time window for any request r_i issued by traveler i is defined by an earliest pick-up time e_{r_i} and a latest pick-up time l_{r_i}. The earliest pick-up time e_{r_i} is set to the estimated arrival time of i at the pick-up location associated with the request. The latest pick-up time l_{r_i} is defined similarly as in chapter 4, based on the cost of the best alternative that does not use AV. Equation 5.8 provides the expression of l_{r_i}.

$$l_{r_i} = e_{r_i} + \frac{1}{\eta_{i,\mathrm{A}}^{m^*}} \min \left(t_{\mathrm{wait}}^{\mathrm{A,max}}, \frac{\min_{m \in M_i \setminus \{\mathrm{A,AP}\}} C_i^m - C_i^{m^*}}{\beta_i^{\mathrm{wait}}} \right) \tag{5.8}$$

where m^* is the mode chosen by i containing $\eta_{i,\mathrm{A}}^{m*}$ AV legs, $t_{\mathrm{wait}}^{\mathrm{A,max}}$ is an upper bound for the cumulative waiting time for AMoD over the whole trip. The difference between l_{r_i} and e_{r_i} is the same for all AV legs composing the optimal route for mode m^*. Note that $\min_{m \in M_i \setminus \{\mathrm{A,AP}\}} C_i^m - C_i^{m^*} \geq 0$ for the the deterministic mode choice model.

Time windows are constraints for the dispatcher of AMoD: a match is proposed to a request only when the estimated pick-up time by an AV verifies the request's time window. Assuming a null pick-up time, we consider the maximum potential of AMoD demand, with potential requests with narrow time windows. In this way, we can capture the choices made by the AMoD dispatcher given its profit-oriented objective: which requests it will prioritize against the others.

When a user's request is declined by the AMoD dispatcher, the optimal routes and generalized travel costs of all modes $m \in M_i \setminus \{A, AP\}$ are recomputed, and the travel decision model is applied on this subset of M_i. The denial can happen while the traveler has not left home yet. In this case, the route's origin corresponds to the user's origin. When the denial happens while the traveler has left home, the route's origin is the last node of the user's current leg.

The walking times $t_{i,\mathrm{walk}}^m$ are computed based on a fixed walking speed v_{walk}.

In this study, we do not compute a user equilibrium to keep the simulation time acceptable and amenable to pricing scheme optimization. To evaluate the quality of the mode and

route choices, we track the difference between the estimated and realized travel cost for each user.

5.4.2 Mobility services step

The same procedure is applied for each mobility service defined in the simulation. It contains two main steps: maintenance and matching. The maintenance relates to vehicles creation and deletion. The matching is in charge of modifying vehicles' plans. For our problem, three mobility services are defined: personal cars, public transportation, and AMoD ride-hailing service. Each has its own maintenance and matching routine.

5.4.2.1 Personal cars and public transportation

Personal cars and public transportation are available mobility services in MnMS. Their maintenance and matching routine are generic.

For personal cars, the procedure is quite simple. The matching consists in creating a car at the user's origin and assigning it a plan containing a pick-up activity and a serving activity without delay. The pick-up activity has an empty route while the serving activity contains the car leg of the user's route. In this way, the user immediately departs from her origin on board her car at the next Vehicles Flow step. The maintenance consists in deleting all cars that have finished their plan, i.e., have reached the end node of their serving activity route.

A user requests a ride to public transportation when she arrives at her boarding stop by specifying her desired alight stop. The matching consists in identifying the next vehicle of the public transportation line chosen by the user to arrive at the boarding stop. Once identified, the user is inserted into this vehicle's plan. The maintenance of public transportation consists in creating the vehicles at the first stop of each line following the line timetables provided in the inputs of the simulation. Additionally, it deletes all vehicles that arrive at their terminus stop.

Note that request cancellation is not possible for personal cars and public transportation.

5.4.2.2 AMoD

The default AMoD ride-hailing mobility service in MnMS was not matching the requirements of this study and significant extensions were needed.

The matching step implements a revised version of the batch-matching over horizon policy described in chapter 4 called iterative batch-matching over horizon. It is executed every dt_{match}. In the following, we describe this revised algorithm.

In this section, r and r_i are equivalent notations. Let us note R_{O}, the set of open AMoD requests at t. A request is *open* when it has been issued and not yet matched or declined. At the initialization of the simulation, R_{O} is empty. In accordance with the simulation flowchart presented in Figure 5.2, the AMoD operator has received all requests of users departing within $[t + H, t + H + dt_{\text{aff}}]$ at t. Hence, one can gather at t all requests $r \in R_{\text{O}}$ for which $e_r \leq t + H$ in a set $R_{\text{O,H}}$. V_{O} designates the set of all available vehicles. A vehicle is *available*

when it has no more than K requests to serve in its plan. Initially, $V_{\mathrm{O}} = V$ where V gathers all AVs of the fleet. The plan of $v \in V_{\mathrm{O}}$ is noted P_{R_v} where R_v is the set of requests that have been matched with v.

The utility $u_{v,r}$ of a match between $r \in R_{\mathrm{O}}$ and $v \in V_{\mathrm{O}}$ is defined as the profit per time unit:

$$u_{v,r} = \frac{\mathrm{profit}(P^*_{R_v \cup \{r\}}) - \mathrm{profit}(P_{R_v})}{t^{\mathrm{end}}_{P^*_{R_v \cup \{r\}}} - t^{\mathrm{end}}_{P_{R_v}}} \tag{5.9}$$

where $t^{\mathrm{end}}_{P_{R_v}}$ is the estimated end time of plan P_{R_v}, $P^*_{R_v \cup \{r\}}$ corresponds to P_{R_v} incremented with the pick-up and serving activities associated with r (inserted at the end), the profit of a plan takes into account the gross income of its serving activities and the operation cost of its pick-up and serving activities. The operation cost of AVs per distance unit is defined by $c^{\mathrm{A}}_{\mathrm{d}}$ parameter.

Algorithm 3 describes the matching step for AMoD service. K_{match} is a parameter of the algorithm that specifies the number of times the maximum cumulative utility matching problem should be solved at most. This problem is the same as in section 4.3.2.4. Such a boundary limits the time spent in the matching step. The condition on $u_{\max}$ plays the same role: if no more profitable requests are in R_{O}, the matching procedure can stop. Condition on step 11 of the algorithm has been added to prevent having users with two different requests within the horizon, one matched and the other declined. The users that have not been matched at the end of the algorithm are considered refused by the service and will go through the diversion procedure described above during the Users step.

When an AV executes its plan, it may arrive ahead of e_r at the request's pick-up location, or the user may arrive later than expected due to delays in the previous legs of her path. If so, the AV waits for the user, delaying its serving activity. Requests' time windows are updated each dt_{match} according to users' progress to prevent cascading delays.

There is no maintenance for this service, as AVs are created during the initialization of the simulation, at the initial positions provided in the inputs, and never deleted.

5.4.3 Vehicles Flow step

In the Vehicles Flow step, vehicles of all types (cars, AVs, buses, metros, and trains) execute their plan and are moved on the network. While metros and trains run on dedicated physical edges at constant speeds, all other vehicle types contribute to the road traffic. MnMS is founded on the trip-based Macroscopic Fundamental Diagram (MFD) framework to describe vehicles motion and traffic dynamics. The MFD consists in a relationship between aggregated traffic variables in a cluster of network links (called a reservoir) within a certain period of time. The speed-MFD links the mean speed of each vehicle type in a reservoir with the accumulations of all vehicle types in this reservoir. Only running vehicles are counted in the accumulation. The trip-based approach takes into account the evolution of accumulation during a trip duration. All vehicles of the same type running within the same reservoir during a dt_{flow} time step are assumed to move at the same speed. The event-based resolution method presented in the chapter 2 of Mariotte (2018) has been adapted into a time-stepped one to keep consistency with the rest of the simulation process. In the default version of

Algorithm 3: Iterative batch-matching over horizon algorithm

if $t = kdt_{\text{match}},\ k \in \mathbb{N}$ **then**
  $V_{\text{O}} \leftarrow \{v \in V,\ |R_v| <= K\}$;
  $R_{\text{O,H}} \leftarrow \{r \in R_{\text{O}},\ e_r \leq t + H\}$;
  iter $\leftarrow 1$;
  $u_{\max} \leftarrow 1$;
  while $|V_{\text{O}}| > 0$ ***and*** $|R_{\text{O,H}}| > 0$ ***and*** iter $\leq K_{\text{match}}$ ***and*** $u_{\max} > 0$ **do**
    Calculate $u_{v,r},\ \forall r \in R_{\text{O}},\ \forall v \in V_{\text{O}}$;
    $u_{\max} \leftarrow \max_{v,r} u_{v,r}$;
    Solve the maximum cumulative utility matching problem;
    for (v^*, r^*) *in the solution matches* **do**
      if *user i has no other requests in* $R_{\text{O,H}}$ *than* r^* ***or*** *all requests of i in* $R_{\text{O,H}}$ *are in the solution matches* **then**
        $P_{R_v} \leftarrow P^*_{R_v \cup \{r^*\}}$ (match);
        Remove r^* from $R_{\text{O,H}}$;
        if $|R_v| > K$ **then**
          Remove v^* from V_{O};
    iter $\leftarrow$ iter $+ 1$;
  Set state of all users with a request $r \in R_{O,H}$ to *refused*;
  $R_O \leftarrow \{\}$;

MnMS vehicular motion, a reservoir is supposed to have no restriction on its inflow and outflow, i.e., there is no queue to access the reservoir or leave it toward another one. This simple approach is valid (remains stable) for our scenarios since the system remains undersaturated, i.e., accumulations of vehicle types never overcome the critical values. To tackle highly congested scenarios, MnMS includes more advanced boundary management principles that are not considered here.

5.5 Multi-objective optimization approach

5.5.1 Choice for an efficient multi-objective resolution approach

To deal with multi-objective optimization problems with black-box objective functions, a wide range of methods have been reported in the literature. Multi-Objective Evolutionary Algorithms (MOEAs) have become popular since they can handle different kinds of decision variables (binary, integer, real, mixed) and do not assume any convexity, differentiability, or linearity of the objective functions involved. Pareto-archived Evolutionary Strategy (PAES, Knowles & Corne (1999)), Strength-Pareto Evolutionary Algorithm (SPEA, Zitzler (1999))and Non-dominated Sorting Genetic Algorithm (NSGA, NSGA-II, Deb *et al.* (2002)) can be cited. However, as population-based algorithms generally require a large number of evaluations, the approach fails on problems with black-box objective functions that are expensive to evaluate.

Model-assisted Evolutionary Strategies (MAESs) try to overcome this limitation. They use a surrogate model that is relatively fast to be trained on a set of evaluated points for each objective (and eventually each constraint) to select a small set of promising points in the new generation that are worth being evaluated and hence, speed up the search. Emmerich *et al.* (2006) compares several criteria for identifying promising points based on the generalization of improvement, probability of improvement, and expected improvement in the multi-objective case. Shinkyu Jeong & Obayashi (2005) (Multi-EGO) evaluates the fitness values of each individual of the population as the Expected Improvements of objective functions from Kriging models. Chugh *et al.* (2018) (K-RVEA) uses Angle Penalized Distance (ADP) criterion, which is better to balance convergence and diversity of the population as it relies on a set of reference vectors that partition the objective space into subspaces where selection of individuals is performed independently.

If MAESs are good for deriving a large pool of optimal solutions, population-based methods are better suited when each generation can be treated as a batch, i.e., when parallelization of evaluations is possible. When finding solutions at the Pareto Front rapidly is preferred over discovering the complete Pareto Front, Multi-Objective Bayesian Optimization (MOBO) can be used Low *et al.* (2022). It is the direct extension of single-objective Bayesian Optimization to solve multi-objective problems, where a separate surrogate model is built for each objective. Many MOBO algorithms have been proposed in the literature. Aggregation-based MOBO emancipates from the multi-objective context by applying BO on a weighted sum of objective functions with the weights randomly varying from one iteration to another to cover the whole Pareto Front, as in Knowles (2006) (ParEGO). On the contrary, dominance-based MOBO considers separate objective functions and relies on the Pareto dominance definition. It defines new infill sampling criteria (or acquisition functions) that guide the subsequent sampling toward the point most likely to improve the current Pareto Front. Several infill criteria have been tested: expected improvement and probability of improvement with respect to multiple objectives (Keane, 2006), expected hypervolume improvement (K-RVEA, Hebbal *et al.* (2019)), hypervolume contribution of Lower Confidence Bound (SMS-EGO, Ponweiser *et al.* (2008)), stepwise uncertainty reduction (SUR, Picheny (2015)).

MOBO is the ideal approach for this study since our simulation model already relies on the available machine cores for some sub-tasks (e.g., the AMoD matching procedure and the update of links costs in the multimodal graph).

5.5.2 MOBO

Among the state-of-the-art MOBO algorithms, Galuzio *et al.* (2020) have introduced a simple yet efficient dominance-based MOBO algorithm called MOBOpt. The algorithm uses a sharp sampling strategy to maximize the diversity of evaluated points (in both search and objective spaces) chosen from a surrogate-based estimation of the Pareto Front. Algorithm 4 presents the revised version of the original MOBOpt[2] we used for this study. In the following, we describe the algorithm steps and provide implementation details.

Let us note K_{init}, the number of initial points evaluated through simulation at step 2 of

[2]https://github.com/ppgaluzio/MOBOpt

the algorithm, and q, the total number of evaluations done so far. In practice, the K_{init} points are chosen with a Latin Hypercube routine. $\mathcal{D}_{f_i}^{(q)} = \{x^{(l)}, f_i(x^{(l)})\}_{l=1,...,q}$ where $x^{(l)} \in \Omega$ designates the l^{th} point at which the objective functions have been evaluated. $\Phi^{(q)}$ and $\Psi^{(q)}$ respectively designate the observed PF and PS after q simulations. The algorithm aims at minimizing the difference between these sets and the actual PF and PS.

Based on $\mathcal{D}_{f_i}^{(q)}$, a surrogate model for each objective function is built at step 5 of the algorithm. Several types of surrogate models can be chosen, such as Polynomials, Neural Networks, or Gaussian Processes. The latter type has the advantage of including a measure of the uncertainty of the function estimation at each point of the definition domain. This interesting feature allows to define acquisition functions that efficiently balance exploitation and exploration in the context of BO. In MOBO, Gaussian Processes are also widely used as surrogate models. $\mathcal{GP}_i(\mu, k|\mathcal{D}_{f_i}^{(q)}, x)$ designates the Gaussian Process of mean $\mu : \Omega \subset \mathbb{R}^{|\mathscr{P}|} \mapsto \mathbb{R}$ and kernel function $k : \Omega^2 \mapsto \mathbb{R}$ inferred from the q first evaluations of f_i.

$$\mathcal{GP}_i(\mu, k|\mathcal{D}_{f_i}^{(q)}, x) = \mathcal{N}(\hat{\mu}_{f_i}^{(q)}(x), [\hat{\sigma}_{f_i}^{(q)}(x)]^2) \tag{5.10}$$

where $\hat{\mu}_{f_i}^{(q)}(x)$ estimates $f_i(x)$ and $\hat{\sigma}_{f_i}^{(q)}(x)$ estimates the prediction error made by the surrogate model at $x \in \Omega$.

The Gaussian Process provides a distribution over the functions consistent with the observed data and the constraints defined by μ (the relevant functions should produce outputs with a mean μ) and k (which describes the smoothness of the searched functions). It defines a prior, which is converted into a posterior once new evaluations of f_i are added to $\mathcal{D}_{f_i}^{(q)}$. The mean and standard deviation of the surrogate model at any $x \in \Omega$ is given by:

$$\hat{\mu}_{f_i}^{(q)}(x) = \mu(x) + K^{(q)}(x)^{\mathrm{T}} \left[K^{(q)}\right]^{-1} \left(F_i^{(q)} - \mu^{(q)}\right) \tag{5.11}$$

$$\hat{\sigma}_{f_i}^{(q)}(x) = \left(k(x,x) - K^{(q)}(x)^{\mathrm{T}} \left[K^{(q)}\right]^{-1} K^{(q)}(x)\right)^{1/2} \tag{5.12}$$

where $K^{(q)}(x)^{\mathrm{T}}$ is the covariance vector obtained by applying the kernel function to the observed data and x ($K^{(q)}(x)^{\mathrm{T}} = [k(x^{(1)}, x), ..., k(x^{(q)}, x)]$), $K^{(q)}$ is the covariance matrix of observed data ($K^{(q)} = [k(x^{(l_1)}, x^{(l_2)})]_{l_1,l_2=1,...,q}$), $F_i^{(q)}$ is the vector of the objective function values at observed points ($F_i^{(q)} = [f_i(x^{(1)}), ..., f_i(x^{(q)})]^{\mathrm{T}}$), and $\mu^{(q)}$ is the mean vector at observed points ($\mu^{(q)} = [\mu(x^{(1)}), ..., \mu(x^{(q)})]^{\mathrm{T}}$). Note that in practice, when no data is available, the mean function μ is null.

Several kernel functions exist and can be combined to measure the similarity between two points. This study uses the sum of a Matérn 5/2 and a white kernel. The white kernel allows to deal with noisy objective functions. Even if our simulation model is deterministic, it enables us to handle local variations in certain of our objective functions, as explained in section 5.7.2.2.

$$k(x, x') = \mathcal{C}_{5/2}(x, x') + \sigma^2 I(x, x') \tag{5.13}$$

$$\mathcal{C}_{5/2}(x,x') = \theta_0 \prod_{j=1}^{|\mathscr{P}|} \left(1 + \frac{\sqrt{5}}{\theta_j}|x_j - x'_j| + \frac{5}{3\theta_j}|x_j - x'_j|^2\right) \exp\left(-\frac{\sqrt{5}}{\theta_j}|x_j - x'_j|\right) \tag{5.14}$$

where $I(x,x') = 1$ if $x = x'$ and 0 otherwise, σ^2 is the noise level associated with the objective function (it is the variance of the noise which is considered to be independently and identically normally-distributed), x_j designates the j^{th} component of vector $x \in \Omega$, and θ_0, ..., $\theta_{|\mathscr{P}|}$ are the typical length scales in each dimension of Ω (the kernel is anisotropic).

The resulting kernel is characterized by $|\mathscr{P}| + 2$ hyperparameters optimized with the L-BFGS-B algorithm to maximize the log-likelihood of observed data. The *scikit-learn* python module is used to proceed to Gaussian processes fitting and prediction.

At step 6 of the algorithm, the fitted Gaussian processes are used to compute the estimated PF noted $\hat{\Phi}^{(q)}$ and PS noted $\hat{\Psi}^{(q)}$ regarding the q evaluated points. The initial problem of minimizing the expensive-to-evaluate f_i functions is transformed into minimizing the fast-to-evaluate $\hat{\mu}_{f_i}^{(q)}$ functions. In MOBOpt, this step is achieved with the NSGA-II algorithm. This genetic algorithm applies elitism by selecting individuals with a crowded comparison operator defined by:

$$x \prec x' \iff \big(\text{rank}(x) < \text{rank}(x')\big) \text{ or } \big(\text{rank}(x) = \text{rank}(x') \text{ and } \text{dist}(x) > \text{dist}(x')\big) \tag{5.15}$$

where the rank refers to the non-dominance rank (rank 1 gathers non-dominated solutions, rank 2 collects all solutions dominated by exactly one other solution, etc.) and is obtained with a fast non-dominated sorting algorithm, dist estimates the distance of a solution to its neighboring solutions in the objective space. A higher distance is preferred to guide the selection toward a uniformly spread-out PF (Deb *et al.*, 2002).

Other efficient algorithms could handle this step. For problems with a few objective functions, selecting the non-dominated points from a large population of evenly distributed points can be sufficient. However, NSGA-II has the advantage of producing the desired number of points in $\hat{\Psi}^{(q)}$ which is of interest for step 7.

Step 7 of the algorithm exploits the information provided by surrogate models. $\hat{\Phi}^{(q)}$ is assumed to be a good approximation of the real PF. Hence, points in $\hat{\Psi}^{(q)}$ are designated candidates for the next evaluation. One point belonging to the estimated PS is selected according to:

$$x^{(q+1)} = \underset{x \in \hat{\Psi}^{(q)}}{\arg\max} \left[w \left(\frac{\delta_{\Omega}^{(q)}(x) - \delta_{\Omega,\mu}^{(q)}}{\delta_{\Omega,\sigma}^{(q)}} \right) + (1-w) \left(\frac{\delta_{f}^{(q)}(x) - \delta_{f,\mu}^{(q)}}{\delta_{f,\sigma}^{(q)}} \right) \right] \tag{5.16}$$

where $\delta_{\Omega}^{(q)}(x)$ is the least Euclidian distance from $x \in \hat{\Psi}^{(q)}$ to all points in $\{x^{(1)}, ..., x^{(q)}\}$, $\delta_{\Omega,\mu}^{(q)}$ (resp. $\delta_{\Omega,\sigma}^{(q)}$) is the mean (resp. standard deviation) of $\{\delta_{\Omega}^{(q)}(x)\}_{x \in \hat{\Psi}^{(q)}}$, $\delta_{f}^{(q)}(x)$ is the least Euclidian distance from $[\hat{\mu}_{f_i}^{(q)}(x)]_{i \in [\![1,n]\!]}$ to all points in $\{f^{(1)}, ..., f^{(q)}\}$ ($f^{(l)} = [f_i(x^{(l)})]_{i \in [\![1,n]\!]}$), $\delta_{f,\mu}^{(q)}$ (resp. $\delta_{f,\sigma}^{(q)}$) is the mean (resp. standard deviation) of $\{\delta_{f}^{(q)}(x)\}_{x \in \hat{\Psi}^{(q)}}$, w is a parameter of the algorithm.

This criterion selects the point of the estimated PF that is the farthest away from all previously evaluated points in the search and the objective spaces. Parameter w balances the weight of the search and objective spaces. When w=1, the sampling method focuses on PS densification. When w=0 it focuses on PF densification. In practice, we set w=0.5.

Since steps 6-7 proceed to the exploitation of the available data, step 8 introduces some exploration in the algorithm. If the original MOBOpt does not rely on $\hat{\sigma}_{f_i}^{(q)}$ to explore the search space, we introduce a new exploration strategy taking advantage of this knowledge. Let $\tilde{x}^{(q+1)}$ designate a mutated $x^{(q+1)}$ where all components are preserved except one randomly chosen. This chosen component can be set to any value within the limits specified by Ω. The set of all possible mutated $x^{(q+1)}$ is noted $\tilde{\Omega}$. The choice for $\tilde{x}^{(q+1)}$ is given by equation 5.17. It selects the point for which the surrogate models make the largest errors. The mutation is applied with a certain probability π.

$$\tilde{x}^{(q+1)} = \underset{x \in \tilde{\Omega}}{\arg\max} \sum_{i=1}^{n} \left(\frac{\hat{\sigma}_{f_i}^{(q)}(x) - \min_{x' \in \tilde{\Omega}} \hat{\sigma}_{f_i}^{(q)}(x')}{\max_{x' \in \tilde{\Omega}} \hat{\sigma}_{f_i}^{(q)}(x') - \min_{x' \in \tilde{\Omega}} \hat{\sigma}_{f_i}^{(q)}(x')} \right) \tag{5.17}$$

Finally, $f_i(x^{(q+1)})$ is evaluated through simulation and the observed PF and PS are updated by taking into account this new data point.

Algorithm 4: Revised MOBOpt algorithm

1 $q \leftarrow K_{\text{init}}$;
2 Initialize $\mathcal{D}_{f_i}^{(q)}, \forall i \in [\![1, n]\!]$;
3 Initialize $\Phi^{(q)}$ and $\Psi^{(q)}$;
4 **while** $q <$ simulation budget **do**
5 Build $\mathcal{GP}_i(\mu, k | \mathcal{D}_{f_i}^{(q)}, x),\ \forall i \in [\![1, n]\!]$;
6 Compute $\hat{\Phi}^{(q)}$ and $\hat{\Psi}^{(q)}$ based on $\hat{\mu}_{f_i}^{(q)}$ $(i \in [\![1, n]\!])$;
7 Choose one point $x^{(q+1)}$ in $\hat{\Psi}^{(q)}$ according to 5.16 ;
8 With probability π, choose a random component of $x^{(q+1)}$ and apply the mutation $x^{(q+1)} \leftarrow \tilde{x}^{(q+1)}$ according to 5.17;
9 $\mathcal{D}_{f_i}^{(q)} \leftarrow \mathcal{D}_{f_i}^{(q)} \cup \{x^{(q+1)}, f_i(x^{(q+1)})\}$;
10 Update $\Phi^{(q)}$ and $\Psi^{(q)}$;
11 $q \leftarrow q + 1$;
12 Return $\Phi^{(q)}$ and $\Psi^{(q)}$;

5.6 Theoretical case study

We have designed a theoretical case study close to the one we use in chapter 4 to develop the approach. Figure 5.3 presents the theoretical urban area considered. As represented in Figure 5.1, the simulator requires four types of inputs related to the roads, the multimodal network, the demand, and the regulation policy.

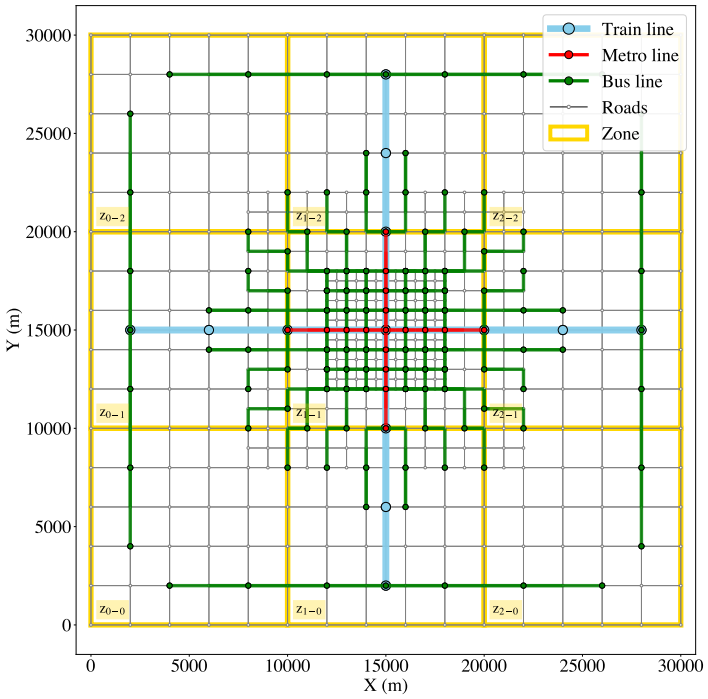


FIGURE 5.3 – *Theoretical urban area.*

5.6.1 Roads

Regarding roads, we consider the same nested Manhattan network as in the previous chapter, with three mesh sizes: the smallest corresponds to the city's dense center, the intermediate one to the suburbs, and the largest to the extended suburbs and close rural area surrounding the city. On top of this network, two railways and two tunnels are defined for the North $\longleftrightarrow$ South and the East $\longleftrightarrow$ West train and metro lines. To fasten the matching process of AMoD service, especially the computation of the utilities for each (v,r) pair, the shortest paths (in terms of distance) between each node of the roads are pre-computed.

As the study case extends on several types of urban areas, the roads sections have been grouped into nine zones as represented in Figure 5.3. Each zone z coincides with one reservoir and is characterized by the following parameters:

- L_z for the total length of roads within z
- N_z for the maximum number of running vehicles within z
- V_z for the free-flow speed in z
- N_z^{c1}, N_z^{c2}, V_z^{c1}, and V_z^{c2} for the critical accumulations and speeds defining the discontinuity points of the piece-wise linear mean speed function in z

As our network is virtual, we have chosen reasonable values for these parameters but did not try to calibrate the MFD functions over micro-simulation data. The maximum accumulation N_z is set proportional to L_z. The free-flow speed is set to 15.5m/s in all reservoirs except the central one, where the free-flow speed equals 11.5m/s. Figure 5.4 shows a typical mean speed function for cars, AVs, and buses which share the same physical network.

For now, all vehicles, including cars, AVs, and buses, are considered to have the same impact on traffic with 2D-MFDs. This assumption is acceptable regarding buses since the public transportation is not changing. Regarding AVs, some micro-simulation studies have shown that depending on the penetration rate of AVs, the MFD form evolves (Lu *et al.*, 2020). Moreover, being part of an on-demand service, the pick-ups and drop-offs on the curbside may impact traffic more than personal cars (Paipuri *et al.*, 2021). Improving the traffic model using multi-reservoir with congestion propagation and calibrated MFD functions is kept for future extension of this work.

Train and metro run on dedicated physical networks. v_{train} and v_{metro} give their respective fixed commercial speeds.

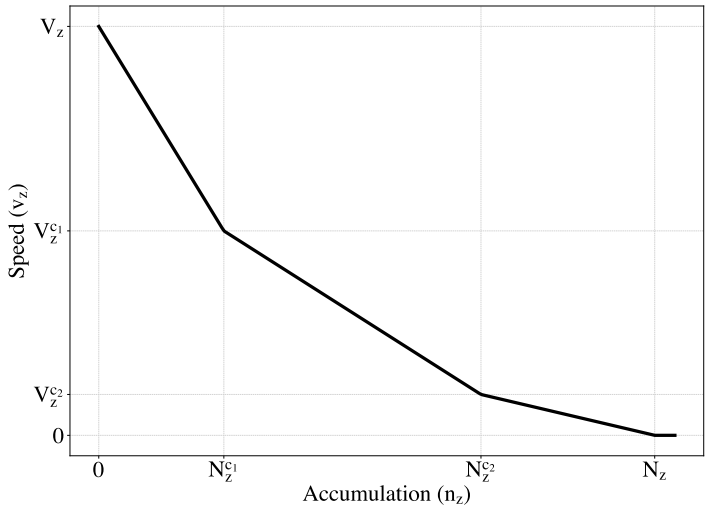


FIGURE 5.4 – *Mean speed over accumulation of vehicles in zone z.*

5.6.2 Multimodal network

As represented in Figure 5.3, the public transportation network contains four train, four metro, and forty bus lines. Each line has a fixed headway for the whole studied period. All bus lines crossing the center have a headway of 10min, while the bus lines that do not cross z_{1-1} (feeding the train and metro lines terminus stops) have a headway of 15min. Train (resp. metro) lines have a headway of 20min (resp. 6min). The public transportation ticket price is 2.5EUR and includes access to all lines mentioned, including the train line. However, riding a train line is charged an additional per distance unit fare of 0.25EUR/km.

Parameters of the AMoD matching procedure described in section 5.4.2.2 are provided in table 5.2. The fleet contains a fixed number of AVs N_{A}. They are initially uniformly distributed over the roads. The base, distance-based, and time-based fares for AMoD service are chosen to ensure the profitability of each service activity for the AMoD operator while maintaining cheap rides for users. The base fare is cheaper than the public transportation ticket price, the distance-based fare ensures 5% benefits for each kilometer served, and the time-based fare is set to 0.1EUR/min.

The simulator requires the definition of travelers' origin and destination nodes to limit the size of the multi-layer graph representing the multimodal network. In our case study, an origin node and a destination node are associated with each node of the roads. In practice, a traveler may depart from an origin located within a road section. However, the assumption is acceptable given the mesh sizes considered in this theoretical city. The set of origins and destinations forms the OD layer.

The multi-layer graph contains four layers here. It includes the car, AMoD, public transportation, and OD layers. Simple connection rules are defined to link layers together and account for the potential walking legs of a traveler's route and flexible intermodality. A connection link is assumed to be traveled by walk. Three connection rules are defined based on two radii: $\rho_{\text{acc/egr}}$ for the maximum access/egress distance by walk, and ρ_{transfer} for the maximum transfer distance by walk.

- The first rule defines the access and egress from the OD layer. Each origin (resp. destination) is linked to all nodes of other layers located within $\rho_{\text{acc/egr}}$ radius by an access (resp. egress) link. After the creation of these connections, for each origin (resp. destination) not connected to any public transportation stop, an access (resp. egress) link is created toward (resp. from) the closest public transportation stop. If several stops are equidistant to the origin (resp. destination), one link is created for each. This second step ensures every traveler has a possible route on **P** mode.

- The second rule defines intra-layer connections. It only concerns the public transportation layer: each stop is linked with all other stops within a ρ_{transfer} radius.

- The last rule defines inter-layer connections. Each node of the car (resp. AMoD, public transportation) layer is linked to all public transportation (resp. public transportation, AMoD) nodes within a ρ_{transfer} radius. Moreover, each car node is linked to each AMoD node located at the same position.

5.6.3 Demand

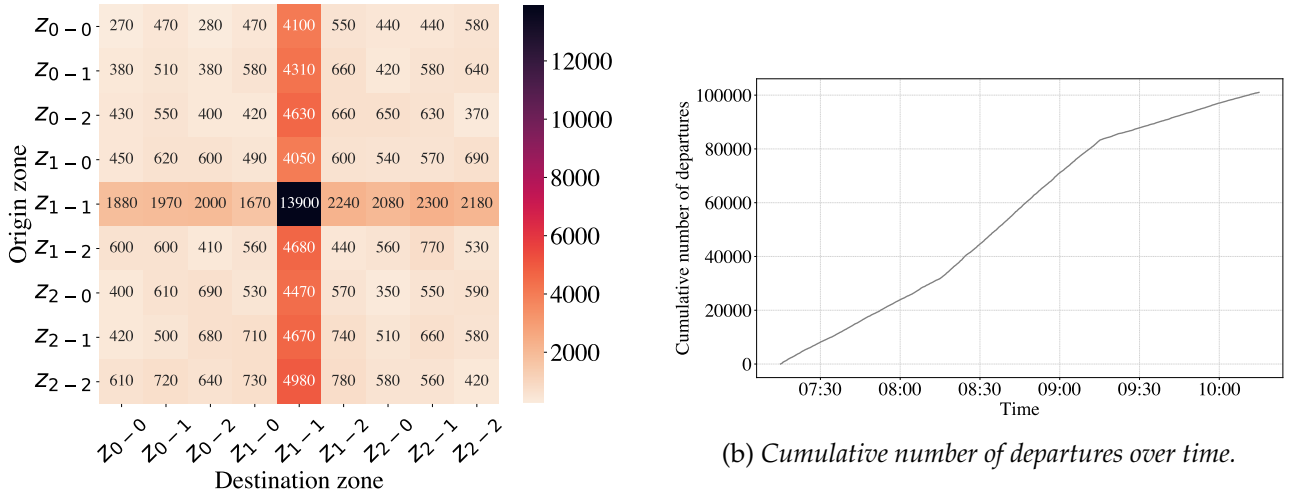
We generate two demand scenarios that mimic the morning commute in our theoretical monocentric city. To keep the simulation time compatible with the optimization of the regulation policy, a scaling factor s is applied. Each simulated traveler, car, and AV represents $s =$ travelers, cars, and AVs. If this downscaling method is widely used in simulation studies [3], this is a limitation of our work and future extension should deal with simulation performance improvement.

The first scenario is focused on non-motorized users, having access to **A**, **P** and **AP** modes only. To keep the traffic dynamics, a background car traffic is defined based on the outputs of a simulation with 14k users scaled with a factor of 10, that have only access to **C**, **P**, and **CP** modes. A set of 6k unscaled non motorized users (representing 4.3% of the total demand) is simulated in this environment.

[3] Diallo *et al.* (2023) uses 5% samples of Lyon and Montréal demands, Ben-Dor *et al.* (2022) uses a 30% sample of Jerusalem demand, Ziemke *et al.* (2019) uses a 10% sample of Berlin demand, Ben-Dor *et al.* (2021) discusses the downscaling validity

The second scenario deals with 10k users scaled with a factor of 10, where θ_C percents (uniformly distributed over the population, independently from their origin and destination) can drive. The other $1 - \theta_C$ percents represent the youngest users without a license and people with disabilities incompatible with the driving activity. Figure 5.5 shows the cumulative departure count curve and the OD matrix for this scenario.

Note that all travelers are considered homogeneous regarding their mode preferences, i.e., they all have the same values of time and transfer penalties (see Table 5.2).



(a) *OD matrix.*

(b) *Cumulative number of departures over time.*

FIGURE 5.5 – *Whole demand scenario (with application of the scaling factor).*

5.6.4 Emissions

In this chapter, CO2 emissions are one of the regulator's main focus. As the public transportation supply does not evolve in our approach, we consider that the related CO2 emissions remain unchanged when the regulation policy evolves. It implies that we neglect the effect of trains, metros, and buses loads on the total emissions of the public transportation system. For the scenario dealing with non-motorized travelers only, the objective is reduced to AVs CO2 emissions equivalence. Indeed, AVs are considered to be electric vehicles. The CO2 emissions per consumed energy are approximated using RTE's 2020 éCO2mix data set[4] for the French energy production. For the complete scenario, cars and AVs emissions are summed.

Cars CO2 emissions factor e_C and AVs energy consumption factor e_A accounts for the effect of the traffic dynamics. They are functions of the reservoir mean speed v_z, which evolves endogenously. The energy consumption curve is borrowed from Lejri *et al.* (2023) for a medium-sized electric vehicle (Renault Zoe 135 ps). It has been built on accurate electric vehicle energetic models (based on VEHLIB software, Vinot *et al.* (2008)) and speed profiles from Hyzem, WLTC and Artemis databases. We include a 17.1% increase in energy consumption due to the weight and operation of sensing, communicating, and computing

[4]See https://www.rte-france.com/eco2mix/les-emissions-de-co2-par-kwh-produit-en-france

subsystems (Gawron *et al.*, 2018), considering that the increase ratio does not depend on the speed. The CO2 emissions curve for personal cars has been built on the COPERT V model, considering a representative pool of personal vehicles for France in 2017. Equation 5.18 expresses the total CO2 emissions for the whole morning commute.

$$E_{\mathrm{CO2}} = \sum_{t=t_{\mathrm{start}}}^{t_{\mathrm{end}}} \sum_{z} \left(e_{\mathrm{C}}(v_z(t)) d_z^{\mathrm{C}}(t) + 1.171\, e_{\mathrm{A}}(v_z(t)) d_z^{\mathrm{A}}(t) \right) \tag{5.18}$$

where $v_z(t)$ is the mean speed in reservoir z between t and $t + \Delta t$, $d_z^{\mathrm{C}}(t)$ (resp. $d_z^{\mathrm{A}}(t)$) is the total distance traveled by cars (resp. AVs -serving and empty-) within z between t and $t + \Delta t$, Δt is an aggregation time step. In practice we choose $\Delta t = dt_{\mathrm{aff}}$.

As we remain macroscopic in traffic dynamics modeling, emissions calculation is also macroscopic. For pollutant emissions, it is known that instantaneous emission models based on speed profiles with fine granularity are required for estimation accuracy. However, we restrain our analysis to CO2 emissions, for which aggregated approaches are relevant. If the scale of application of these models matters (Lejri & Leclercq, 2020), the comparison of scenarios for which emissions have been computed at the same scale remains informative.

Roads			
$V_{z_{1-1}}$	11.5	(m/s)	Free-flow speed in z_{1-1}
$V_{z_{0-0}}$	15.5	(m/s)	Free-flow speed in z_{0-0} (same for all zones except z_{1-1})
dt_{flow}	1	(min)	Flow time step
Multimodal network			
v_{walk}	1.42	(m/s)	Walking speed
v_{train}	18	(m/s)	Commercial speed of train
h_{train}	20	(min)	Headway of train lines
v_{metro}	13	(m/s)	Commercial speed of metro
h_{metro}	6	(m/s)	Headway of metro lines
h_{bus}	10,15	(min)	Headway of bus lines
f_0^{P}	2.5	(EUR)	Public transportation ticket price
f_d^{P}	0.25	(EUR/km)	Additional per distance unit fare for using the train line
N_{A}	600	(AVs)	AMoD fleet size
dt_{match}	5	(min)	AMoD service matching time step
H	15	(min)	Horizon length
K	3	(requests)	Maximum number of requests within the plan of AVs
K_{match}	3	(iterations)	Maximum number of iterations to be achieved in the iterative batch-matching over horizon algorithm per dt_{match}
c_{d}^{A}	0.38	(EUR/km)	Operational cost of an AV per distance unit[1]
f_0^{A}	1	(EUR)	AMoD base fare
f_{d}^{A}	0.4	(EUR/km)	AMoD distance-based fare
f_{t}^{A}	0.1	(EUR/min)	AMoD time-based fare
$\rho_{\text{acc/egr}}$	500	(m)	Maximum access/egress walking distance
ρ_{transfer}	200	(m)	Maximum transfer walking distance
c_{d}^{C}	0.68	(EUR/km)	Operational and ownership cost of a car per distance unit[2]
Demand			
β^{drive}	5.84	(EUR/h)	Value of time for driving activity[3]
β^{rideA}	3.93	(EUR/h)	Value of time for riding AV activity[3]
β^{rideP}	2.02	(EUR/h)	Value of time for riding public transportation activity[3]
β^{walk}	18.89	(EUR/h)	Value of time for walking activity[3]
β^{wait}	11.04	(EUR/h)	Value of time for waiting activity[3]
γ^{transfer}	1.07	(EUR)	Transfer penalty[4]
$t_{\text{wait}}^{\text{A,max}}$	15	(min)	Maximum cumulative waiting time for AMoD over a trip
dt_{aff}	15	(min)	Affectation time step
θ_{C}	88	(%)	Users having access to **C** in the whole demand scenario
s	10	(users)	Scaling factor for the whole demand scenario

TABLE 5.2 – *Case study parameters.*

[1] From Becker *et al.* (2020), assuming a Zurich-like city for a AMoD ride-hailing service

[2] From https://www.automobile-club.org/actualites/la-vie-de-l-aca/budget-de-l-automobiliste-de-l-aca-avril-2020 considering the mean value for the panel of cars evaluated

[3] From Kolarova *et al.* (2018), assuming a middle income class of users and actualized euros

[4] From Frei *et al.* (2017), assuming the value for two transfers multiplied by the value of time of driving activity

5.7 Numerical results

This section investigates several regulation policies and objectives based on the simulation-multi-objective optimization approach presented above.

5.7.1 Taxing cars and door-to-door AV rides

Regulating AVs only in the multimodal system is not relevant. Indeed, when the travel cost by AVs for door-to-door rides increases too much, travelers may prefer to use their cars instead. The modal shift from mode **A** to mode **C** increases the system-wide carbon emissions. To make the regulation on mode **A** relevant, mode **C** should be regulated jointly. The main question tackled in this subsection is how to regulate cars with regard to the regulation of door-to-door AV rides. How much should mode **C** be taxed to make the tax on door-to-door AV rides relevant?

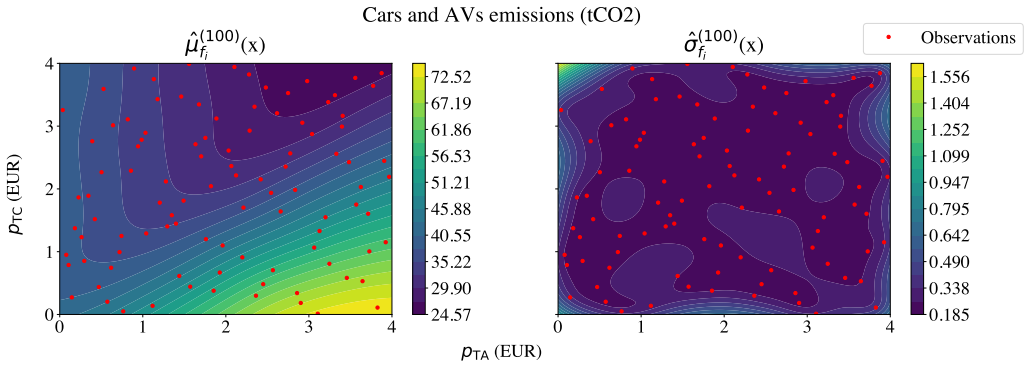
Car trips can be taxed in several ways. The tax can be flat, like a daily congestion charge, and applied to any car trip (including those in connection with a PT station), or kilometric and applied to thermal cars, which impact the most system-wide carbon emissions (similarly to a fuel tax, for example). Here, we investigate flat taxes for door-to-door AV ride (p_{TA}) and cars (p_{TC}). Travelers pay the price p_{TA} to the regulator when they choose **A** mode, and p_{TC} when they choose **C** mode.

A hundred MnMS simulations of the second scenario corresponding to different taxes values within a reasonable search space are launched. By reasonable, we mean a search space revealing some effects on the indicators we track, including emissions. Intuitively, there is a threshold on the car tax above which no one uses this mode anymore, and the tax becomes a dummy variable. The same goes for the tax on door-to-door AV rides. The range [0 EUR, 4 EUR] is chosen for both taxes.

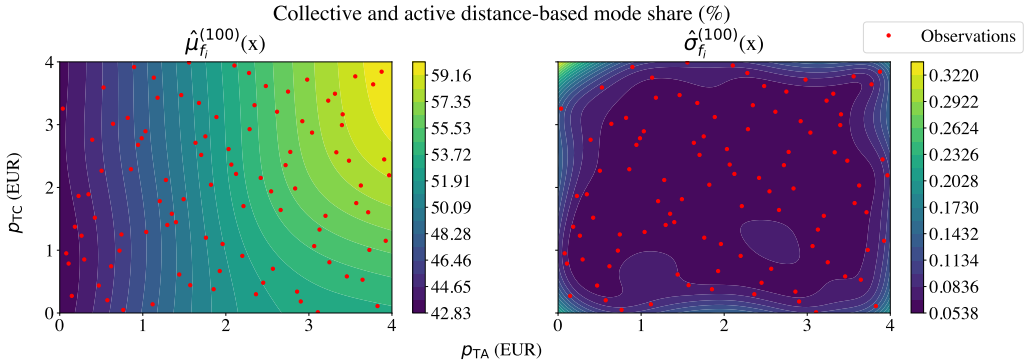
The $\mathcal{D}_{f_i}^{(100)}$ for several indicators f_i are used to fit corresponding surrogate models. We do not enter into the loop of MOBOpt in this subsection but exploit the surrogate models to deduce sensibility tendencies.

Figure 5.6 shows some of the surrogate models. They all produce reasonable errors. The first element to note is that, as expected, increasing p_{TA} with a null car tax increases emissions because of a modal shift from mode **A** to mode **C**. The total distance traveled by cars increases. The modal shift also happens from mode **A** to mode **AP** and **P** as the collective and active distance-based mode share increases (Figure 5.6b), but it remains marginal. With a higher collective and active distance-based mode share, we worsen the carbon footprint because all personal cars are considered thermal and have a much higher carbon emissions factor than electric AVs. The non relevance of p_{TA} maintains till a car tax around 1EUR.

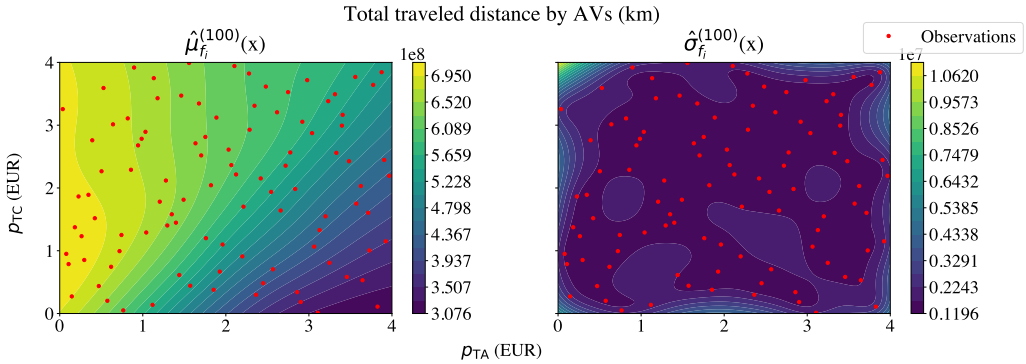
At a medium car tax value (between 1 EUR and 4 EUR), emissions decrease, then increase as p_{TA} grows. The relevancy range of p_{TA} does not cover the whole search interval but widens as the car tax increases.



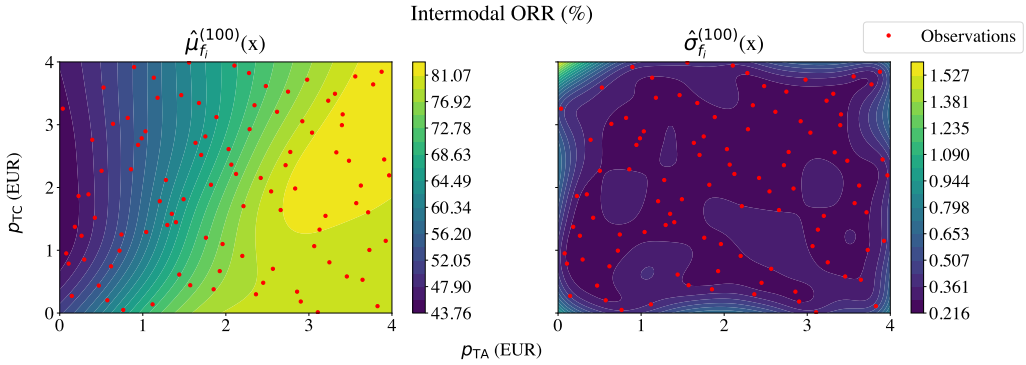
(a) *Surrogate model for cars and AVs emissions.*



(b) *Surrogate model for collective and active distance-based mode share.*



(c) *Surrogate model for the total traveled distance by AVs.*



(d) *Surrogate model for the intermodal Order Response Rate (ORR).*

FIGURE 5.6 – *Surrogate models built on* $\mathcal{D}_{f_i}^{(100)}$ *showing the sensibility of several indicators on* p_{TA} *and* p_{TC} *for the whole demand scenario.*

The entire search space for p_{TA} is relevant from a car tax of 4 EUR. Whatever the car price, the total distance traveled by AVs decreases with p_{TA}. However, in the relevant range of values for p_{TA}, the reduction is smaller (Figure 5.6c). It shows that outside of this range, AMoD loses market share drastically. In contrast, within the relevant range for p_{TA}, i.e., when emissions decrease as p_{TA} grows, AMoD can replace part of its former market share with intermodal requests that are not taxed. The AMoD profit indicator confirms it. Travelers divert from unsatisfied **A** or **AP** mode (they are refused because of a lack of AVs and take their car instead) to **AP** mode. Indeed, increasing the tax on door-to-door AV rides frees some AVs for intermodal rides, which are, on average, smaller than door-to-door trips. A clue for that is the surrogate model for intermodal ORR shown in Figure 5.6d: it increases with p_{TA} whatever p_{TC} but the peak is obtained for medium and high car toll. The tax p_{TA} is definitely relevant for encouraging the synergy between PT and AMoD.

In conclusion, this experiment highlights the upper bound of the relevant search space for p_{TA} depending on the value of p_{TC}. It appears from figure 5.6a that the range of relevancy is above the first bisector. Hence, a possible strategy is setting p_{TC} to the upper bound of the search space for p_{TA}. Another strategy to limit the tax on travelers is to set p_{TC}=p_{TA}. Depending on the regulation policy and the need for more or less income from taxes to pay the subsidies of the scheme, one or the other strategy can be chosen.

5.7.2 Minimizing emissions while financing subsidies with taxes

The money the regulator earns through cars and door-to-door taxes can be re-injected into the system to lower the travel costs of users who opt for collective modes. In this section, we investigate a pricing scheme with two prices: the tax on door-to-door AV rides (p_{TA}) and a subsidy for intermodal AV rides ($p_{TAP} \leq 0$). The risk of providing a subsidy is to induce too much demand coming from the taxed modes. Consequently, the community spends a considerable amount of money to modify travelers' behaviors. Hence, the proper trade-off between the emissions reduction and the cost of the pricing scheme should be found. Additionally, the regulator's aim is not to make some profit with this particular pricing scheme. Then, in this section, we consider minimizing the system-wide carbon emissions while tending to a revenue-neutral scheme.

5.7.2.1 Non-motorized travelers scenario

MOBOpt algorithm is launched with two strategies under a budget of 200 simulations: a large number of initial points followed by a few iterations ($K_{init} = 100$), and a small number of initial points followed by a lot of iterations ($K_{init} = 6$).

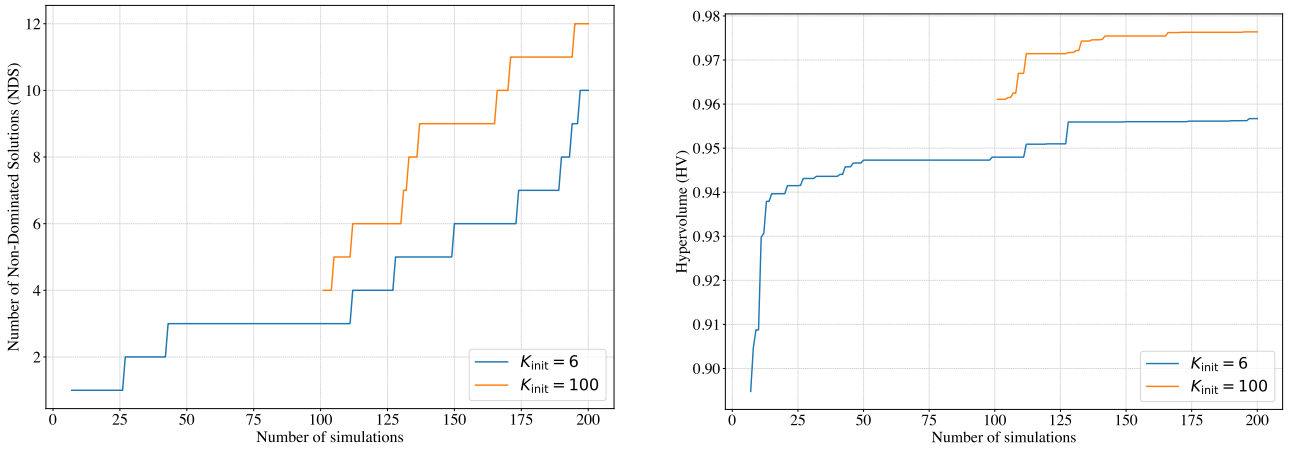
The quality of a Pareto front can be assessed in view of several aspects. The three main aspects are the convergence (how "optimal" the PF is, how close the observed PF is to the real PF, if known), the density which can be sub-categorized into the spread (how large the region that the solution set covers is) and uniformity (how evenly the solutions are spread), and finally the cardinality (how many solutions have been found) (Li *et al.*, 2022). Here, two metrics are tracked to cover these aspects:

- The Number of Non-Dominated Solutions (NDS) metric, which follows how many points from the final observed Pareto set have been found at a certain iteration. A

higher NDS corresponds to a better cardinality.

- The Hypervolume (HV) metric, which follows the volume of the union of the hypercubes determined by each of the solutions and a reference point. The reference point can be chosen as a vector gathering the worst objective functions values. The objectives values are normalized to eliminate the arbitrary scaling of the different objectives. A higher HV corresponds to a better PF in the three aspects.

Figure 5.7 compares the evolution of these quality metrics along the iterations of MOBOpt for both strategies. The $K_{\text{init}} = 100$ strategy reaches higher NDS and normalized HV within the 200 simulations than the $K_{\text{init}} = 6$ strategy. Moreover, HV for $K_{\text{init}} = 6$ never reaches the HV $K_{\text{init}} = 100$ strategy obtains from the 100 initial points. For the rest of the paper, MOBOpt will be launched with a subsequent initial set of points.



(a) *NDS convergence.*

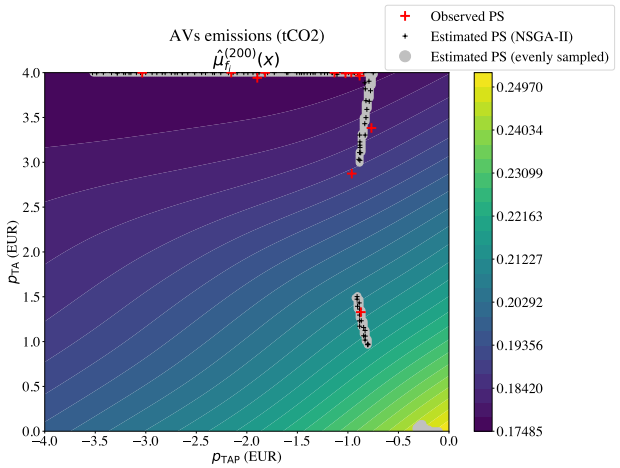
(b) *HV convergence.*

FIGURE 5.7 – *Comparing convergence of MOBOpt algorithm starting with a small (6) and big (100) initial set of points (w=0.5, π=0.1 and linearly decreases over iterations).*

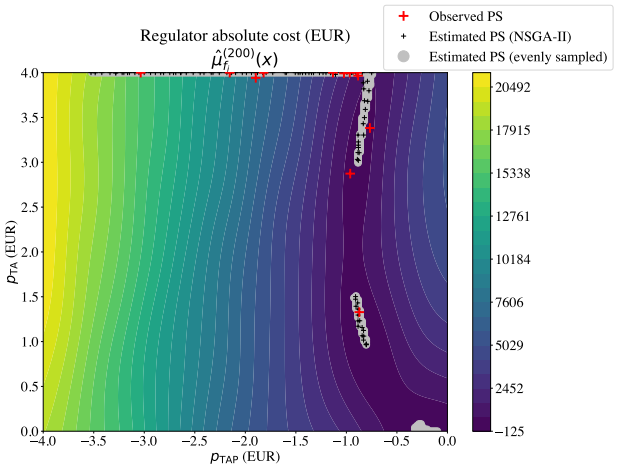
Figure 5.8 presents the estimated and observed PF and PS obtained for the best strategy, namely $K_{\text{init}} = 100$. It also presents an estimated PF and PS obtained via estimating objectives at evenly distributed points using the surrogate models. The estimated and observed PFs suggest that most of the CO2 emissions reductions that can be achieved in the system through the pricing scheme can be reached by maintaining the regulator's monetary balance nearly null. Indeed, the PF profile is sharp, and the normalized HV is close to 1. Spending more money on subsidies than earned from taxes or taking more than subsidizing leads to a marginal reduction of emissions.

Figure 5.8c labels the most promising regulation observed as regulation NM1. It corresponds to a subsidy for intermodal AV rides of 0.88 EUR and a tax on door-to-door AV rides of 3.97 EUR. These prices lead to a reduction of 30% in AVs emissions and costs 982 EUR to the regulator. While the total travel cost for users grows by 4.4%, the increase in the total travel time is marginal (0.04%). The pricing scheme develops the intermodality between AVs and PT. Figure 5.9a shows the modification of the mobility and AMoD service patterns. The number of AV rides increases under regulation, but the percentage of door-to-door rides reduces from 75% to 17%. The median of AV rides distances reduces from 8.5 km to 3 km.

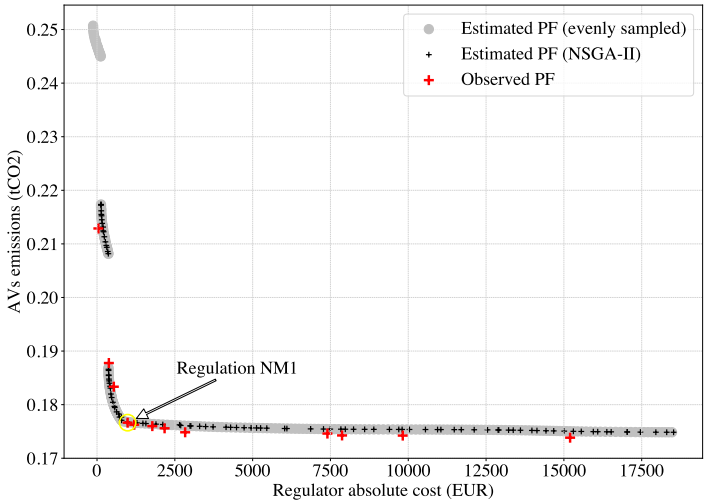
The fleet serves more requests for shorter distances but travels the same number of kilometers empty. The global ORR grows from 77% to 88%. However, the ratio between the empty vehicle kilometers traveled (EVKT) and the total distance traveled by AVs rises from 13% to 19%. As a consequence, the AMoD operator profit reduces by 14%.



(a) *Estimated and observed PS represented on the AVs emissions surrogate model.*



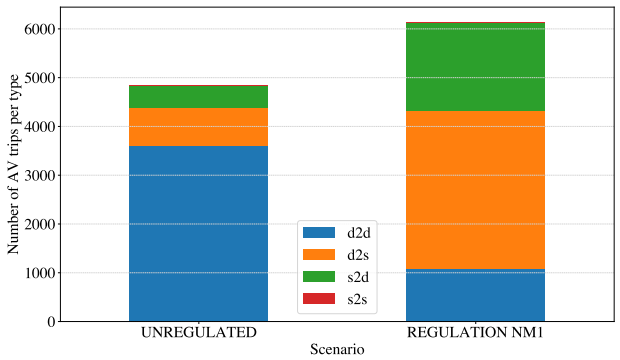
(b) *Estimated and observed PS represented on the regulator absolute cost surrogate model.*



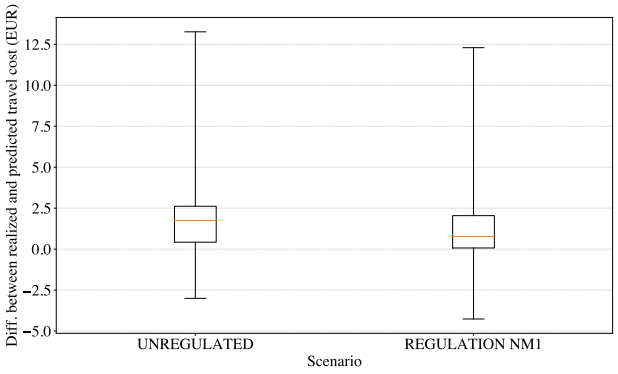
(c) *Estimated and observed PF.*

FIGURE 5.8 – *Estimated and observed PF and PS for AVs emissions and regulator absolute cost objectives on the non-motorized demand scenario.*

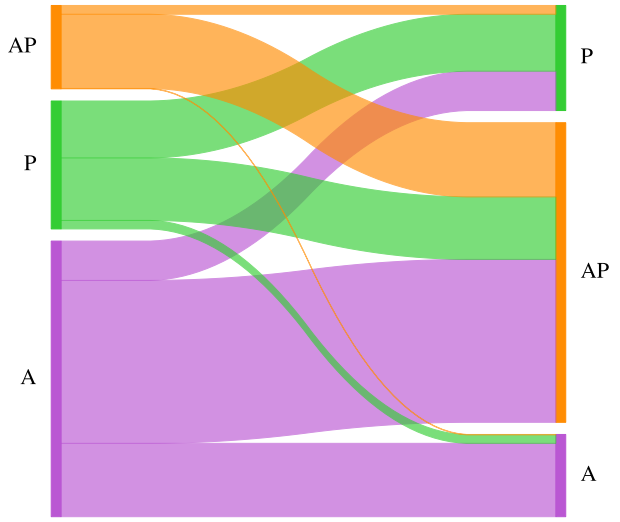
Figure 5.9c shows the modal shift between the unregulated (left) and the regulated (right) scenarios. Note that the subsidy for mode **AP** attracts a substantial proportion of users that used to take **P** mode. These are travelers using AVs to replace either walking for access and egress PT, or a long train leg. The tax on **A** mode makes some travelers diverting from **A** to **P** mode. Finally, as the fleet is less busy under regulation, it allows a small proportion of travelers that used to take **P** mode to divert toward **A** mode. Following these shifts, the total distance traveled by all modes has increased by 3%.



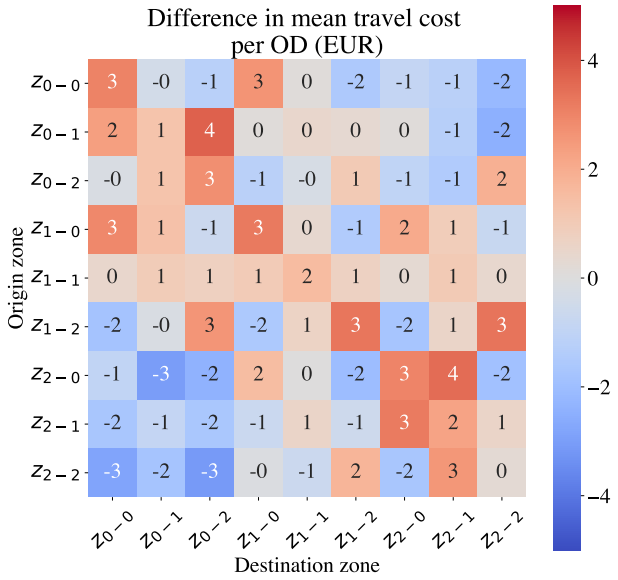
(a) *AV rides count per type, d2d, d2s, s2d and s2s respectively stand for door-to-door, door-to-station, station-to-door, station-to-station.*



(b) *Distribution of the differences between predicted and realized travel cost.*



(c) *Mode shift from the unregulated (left) to the regulation NM1 (right) scenario.*



(d) *Differences in average travel costs per origin zone - destination zone.*

FIGURE 5.9 – *Comparison between unregulated and regulation NM1 scenarios for the case of non-motorized demand.*

Table 5.3 summarizes the changes in the distance traveled per mode. Metro and bus are the PT modes that gain the most market share. In terms of crowding in metro rolling stocks and buses, it corresponds to less than two additional passengers per vehicle on average, given the PT supply defined in the previous section. Walking and train lose ground: it corresponds to travelers shifting from **P** to **AP**. In the suburban zones, using an AV to join one of the bus terminal stations becomes more interesting than walking and using the train, which has an additional kilometric fare compared to the other PT types. Moreover, AV rides being shorter, the ORR in the suburbs is higher in the regulated scenario, so travelers who used to walk to a PT station can now get an AV. For almost all origin zone - destination zone (OzDz) pairs internal to the suburbs, the walking distances fall as shown by Figure 5.10. The mean gap is higher for trips between far suburban zones and reaches 1 km for z_{0-0}-z_{2-2}. The fall in walking distances does not affect all OzDz pairs. Users from or toward the city center walk approximately the same number of kilometers. However, users traveling within the

center walk on average 200 meters more in the regulated scenario. A total number of 206 additional kilometers are walked within the city center. It corresponds to travelers shifting from **A** to **P** or **AP**.

The new mobility pattern leads to better reliability of the predicted travel cost as shown by figure 5.9b. However, Figure 5.9d shows that travel costs changes are not homogeneous over the OzDz pairs. As expected, some users benefit from the pricing scheme (in blue), while others experience higher travel costs (in red). Intra-zonal trips, trips between neighboring suburban zones, and trips from the city center are disadvantaged. On the contrary, trips between far suburban zones tend to have lower travel costs.

Walk	-13.8%
Train	-19.4%
Metro	+41.7%
Bus	+64.4%
AV	-35.8%
Empty AV	+2.6%

TABLE 5.3 – *Changes in the total distances traveled per mode between unregulated and regulation NM1 scenarios for the case of non-motorized demand.*

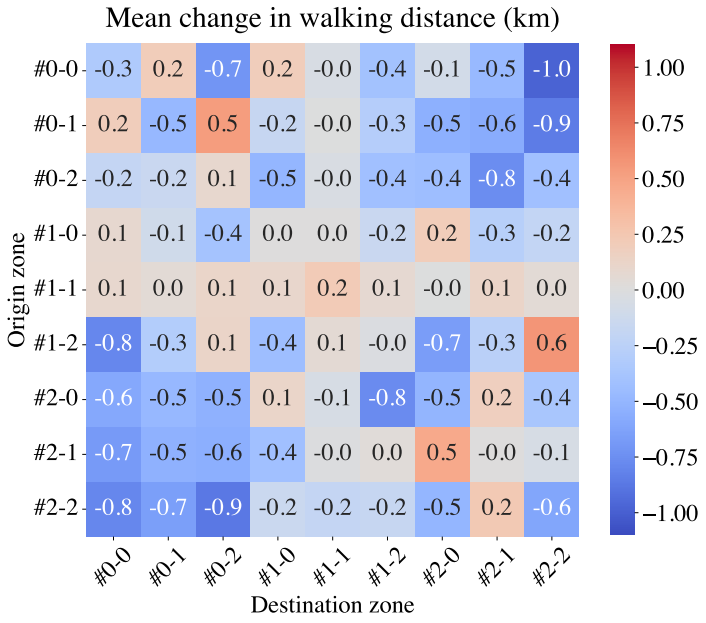


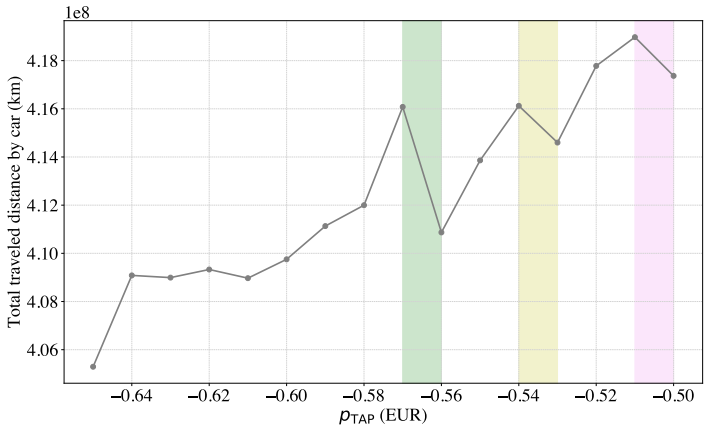
FIGURE 5.10 – *Mean change in walking distances per origin zone - destination zone between unregulated and regulation NM1 scenarios for the case of non-motorized demand.*

5.7.2.2 Whole demand scenario

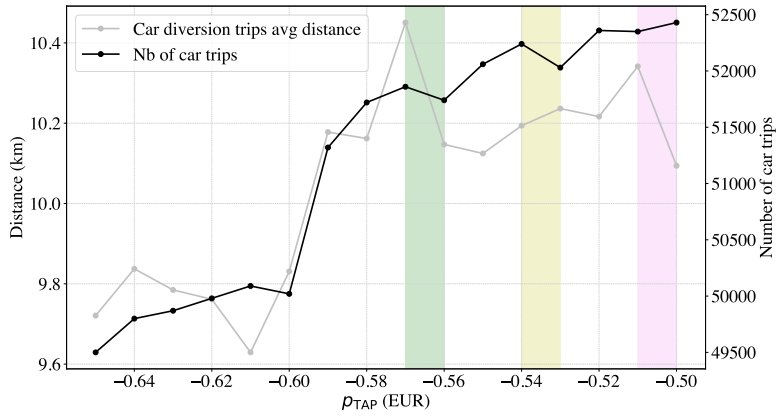
This subsection performs the same investigation on the whole demand scenario. Regarding the results of section 5.7.1, the tax on cars is set equal to the tax on door-to-door AV rides

(p_{TC}=p_{TA}).

When the majority of the demand considered is flexible in terms of mode of travel, cars and AVs emissions objective, which is directly linked to the total distance traveled by cars [5], have local optima. As shown in Figure 5.11a, the total distance traveled by cars globally decreases when the subsidy for intermodal AV rides increases, but local optima exist. It is mainly due to the sensibility of the AMoD dispatching strategy. Depending on the spatial and temporal characteristics of AMoD demand, the configuration can be more or less favorable to travelers with **C** or **CP** as diversion modes when no AV is available, or it is profitable for no AV to be matched with them. Figures 5.11a and 5.11b highlight three local optima.



(a) *Total distance traveled by cars.*



(b) *Average distance of diversion car trips (car trips occurring because AMoD declined the user) and total number of car trips.*

FIGURE 5.11 – *Local optima of the total distance traveled by cars for close p_{TAP} values with fixed p_{TC}=0 EUR and p_{TA}=3 EUR.*

The purple area corresponds to a local increase in the total distance traveled by cars and the average distance of car diversion trips. In contrast, the total number of car trips decreases. AMoD has declined fewer potential drivers, but those who have been declined drive for more kilometers. The yellow area corresponds to the opposite situation: the average distance of car diversion trips increases, but more potential drivers have been declined.

[5] As AVs are considered electric, they have a limited contribution to this objective

The green area corresponds to a subsequent increase in the total distance traveled by cars because more potential drivers have been declined, and the average distance of car diversion trips increases.

Local variations remain small enough compared to the global range of emissions to be ignored for designing the regulation policy. The white kernel allows us to consider local optima as noise. Consequently, the estimation provided by the surrogate model may be relatively far from the actual objective value. Since we are tackling a minimization problem, the observed PF contains points corresponding to local minima. As shown on Figure 5.13c, the estimated PF is above the observed PF.

Despite over-estimating surrogate models, a budget of 200 simulations led to an observed PF with 10 points (Figure 5.12), exhibiting similar results as in the non-motorized demand scenario. Indeed, Figure 5.13c shows that the biggest reduction in CO2 emissions can be achieved by keeping the regulator's monetary balance almost null. Points in the observed PF located further away from a null monetary balance correspond to applying more taxes than providing subsidies and lead to marginal gains in terms of emissions.

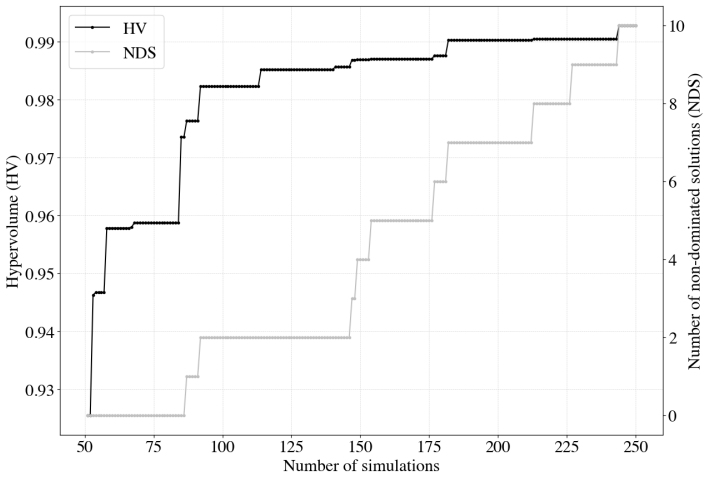
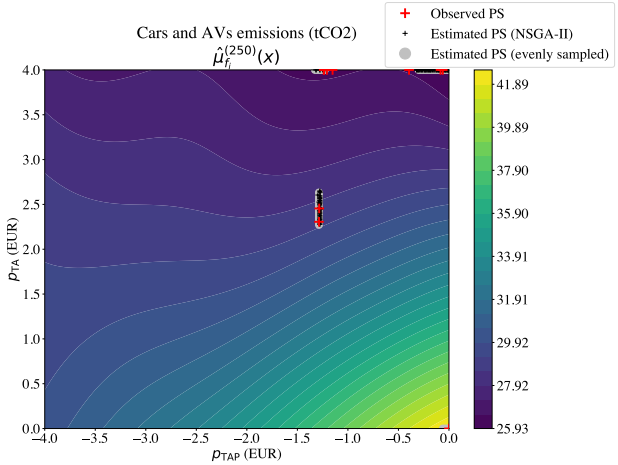
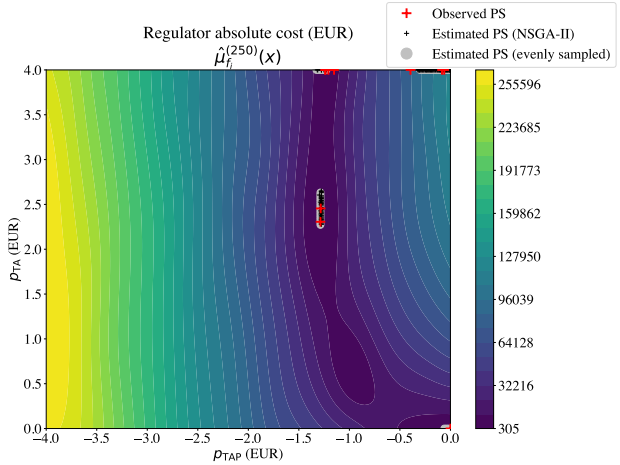


FIGURE 5.12 – *Convergence of MOBOpt algorithm for the whole demand scenario (w=0.5, π=0.7 and linearly decreases over iterations).*

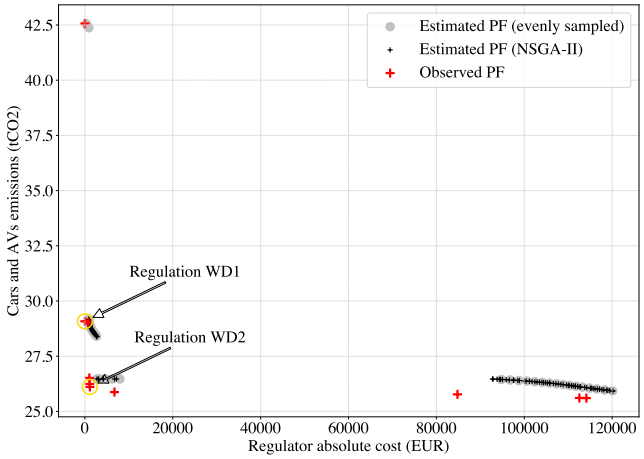
The estimated PF is discontinuous. Unlike the non-motorized demand scenario, discontinuities occur in both "branches" of the PF (decreasing regulator absolute cost for increasing emissions and decreasing emissions for increasing regulator absolute cost). Four areas within Ω emerge from the observed and estimated PS in Figures 5.13a and 5.13b. First, null prices correspond to the highest carbon footprint. Second, a subsidy for intermodal AV rides of 1.25 EUR with a tax on door-to-door AV rides of 2.5 EUR allows to reduce the system's carbon footprint by 32% while maintaining a nearly null monetary balance. Third, increasing the tax to the maximum value of 4 EUR while maintaining the subsidy around 1.25 EUR allows to reduce the emissions by seven additional percent. Fourth, imposing the maximum tax on door-to-door AV rides with a negligible subsidy leads to tremendous income for the regulator. However, emissions reduce by no more than one additional percent. Note that a subsidy higher than 1.5 EUR is irrelevant because it induces too many new requests for intermodal AV rides and is expensive for the community.



(a) *Estimated and observed PS represented on the cars and AVs emissions surrogate model.*



(b) *Estimated and observed PS represented on the regulator absolute cost surrogate model.*



(c) *Estimated and observed PF.*

FIGURE 5.13 – *Estimated and observed PF and PS for cars and AVs emissions and regulator absolute cost objectives on the whole demand scenario.*

Figure 5.13c labels two promising regulations: WD1 and WD2. In the first one, p_{TAP}=-1.28 EUR and p_{TA}=p_{TC}=2.45 EUR. In the second one, p_{TAP}=-1.21 EUR and p_{TA}=4.00 EUR. Table 5.4 compares these two regulation policies with a baseline scenario without AMoD and the unregulated scenario. Regarding the regulator's objectives, both regulations lead to a higher collective and active distance-based mode share than the baseline scenario. The first or second regulation can be favored depending on the strictness of the null monetary balance constraint.

Regarding travelers indicators, the unregulated scenario is preferred to lower the total travel cost, but the baseline scenario guarantees higher reliability of the estimated travel cost despite the higher accumulations and lower speeds in reservoirs due to personal car usage. It is inherent to our choice to consider the maximum potential of demand for AMoD, considering a null predicted pick-up time. If the differences between predicted and realized travel costs remain acceptable, including a convergence loop to compute a day-to-day user equilibrium in the system may slightly worsen the regulator's indicator under regulation.

In terms of AMoD indicators, the unregulated scenario maximizes the AMoD operator's profit while minimizing the ratio between the EVKT and the total distance traveled by AVs.

	No AMoD	**No regulation**	**Regulation WD1**	**Regulation WD2**
		Regulator		
CO2 emissions (t)	121.69	42.57	29.08	26.11
Collective+active distance-based mode share (%)	54.03	42.76	57.75	61.29
Pricing scheme cost (EUR)	-	-	-7.75	-1134
		Travelers		
Total travel cost (kEUR)	1 137.7	1 008.4	1 106.6	1 154.9
3rd quantile of differences between predicted and realized travel costs (EUR)	0.09	1.64	2.19	2.34
		AMoD		
Profit (kEUR)	-	108.9	105.6	103.7
Ratio between EVKT and VKT (%)	-	11.3	17.7	19.4

TABLE 5.4 – *Comparing baseline scenario without AMoD, scenario with AMoD but no regulation, and the two promising regulation policies emerging from the observed PF for the whole demand case.*

5.7.3 Accounting for regulator, AMoD operator and travelers objectives

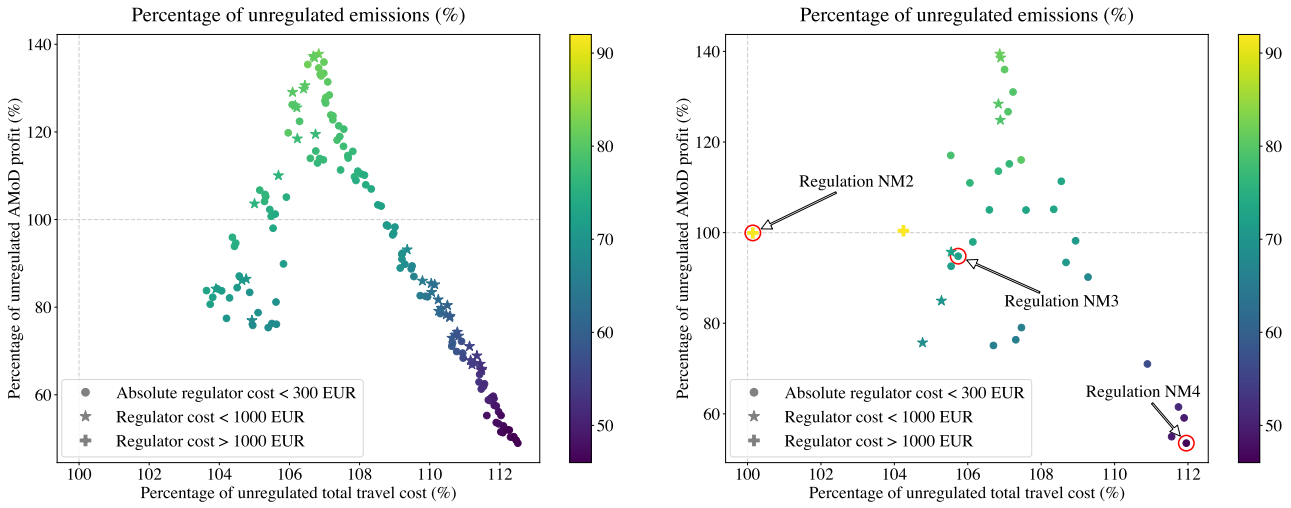
One element highlighted in the previous section is that regulator objectives conflict with travelers and AMoD operator objectives. In the non-motorized demand scenario, the promising regulation leads to the rise of the total travel cost and the reduction of AMoD operator profit. Similarly, on the whole demand scenario, travelers and the AMoD operator prefer the unregulated scenario instead of the promising regulations. Moreover, the prices investigated in the previous section exploit travelers' behavior only. Taxes and subsidies are entirely paid and received by travelers. The behavior of AVs is indirectly affected by the change in the demand pattern for AMoD. In this section, we wonder if we can reach better trade-offs between the regulator, AMoD operator, and travelers objectives by considering taxing and subsidizing travelers on one hand and AVs on the other hand.

The non-motorized demand scenario is investigated with two new prices: a subsidy provided to AVs for serving an intermodal ride p_{AAP} ($\leq$ 0), and a tax imposed to AVs

for serving door-to-door rides p_{AA}. Three objectives are considered in the multi-objective optimization problem. The regulator's objective is to minimize its generalized cost, which is the sum of the pricing scheme's absolute cost and the monetarized emissions. The supposed cost per tCO2 is deduced from the results of section 5.7.2.1 where 76.5 kgCO2 were saved with a pricing scheme absolute cost of 982 EUR. AMoD operator's objective is maximizing its profit. Travelers' objective is to minimize the sum of their travel costs.

A penalty is added to the estimation of objectives in the NSGA-II algorithm to orient MOBOpt search on interesting PF areas. If the regulator's objective estimated by $\hat{\mu}_{f_i}^{(q)}(x)$ is higher than its value evaluated in the unregulated scenario, the penalty is added to the estimation of objectives. The penalty value is chosen large enough to prevent the selection of individuals that do not lead to an improvement of the regulator's objective. Consequently, the estimated PF extends only on relevant areas for this investigation.

Figure 5.14 presents the estimated and observed PFs after 280 simulations, where the observed PF have been filtered to keep only the points verifying a lower generalized cost for the regulator. Figure 5.14a and 5.14b compare the estimated PF with the filtered observed PF where the two components of regulator's generalized cost have been separated. Emissions, AMoD profit, and total travel cost are expressed as percentages of the same quantities in the unregulated scenario.



(a) *Estimated PF.*

(b) *Filtered observed PF.*

FIGURE 5.14 – *Estimated and filtered observed PF for regulator, AMoD operator and travelers objectives on the non-motorized demand scenario.*

The filtered observed PF contains 35 regulation policies improving the regulator's objective. Half of these points lead to a smaller profit for the AMoD operator, but all lead to a higher total travel cost for users. The range of AMoD profit extends between 54% and 139% of the profit achieved in the unregulated scenario. The range of total travel cost extends between 101% and 112% of the total travel cost achieved in the unregulated scenario. Estimated and observed PFs are consistent. However, the evaluated point leading to the second smallest increase in the total travel cost also leads to the second smallest reduction in emissions. A similar point does not exist in the estimated PF.

Regulation NM2 in Figure 5.14b guarantees unchanged AMoD operator profit and total travel cost, a 9% reduction in emissions, but costs 11 kEUR to the regulator. The corresponding pricing scheme is: p_{TAP}=-0.85 EUR, p_{TA}= 1.40 EUR, p_{AAP}= -0.42 EUR, and p_{AA}= 0.57 EUR.

All observed points in PF except the two mentioned above produce reasonable regulator absolute costs. A third noteworthy point is the one providing approximately the same reduction/rise (5.5%) for AMoD profit and total travel cost. It leads to a 28% reduction in emissions with the following prices: p_{TAP}=-0.55 EUR, p_{TA}= 4.0 EUR, p_{AAP}= -0.34 EUR, and p_{AA}= 1.10 EUR.

A fourth noteworthy point is labeled NM4 in Figure 5.14b and corresponds to the highest reduction in emissions (54%) but is to the disadvantage of users and AMoD operator. The corresponding prices are: p_{TAP}=-0.0 EUR, p_{TA}= 2.85 EUR, p_{AAP}= -0.05 EUR, and p_{AA}= 3.79 EUR.

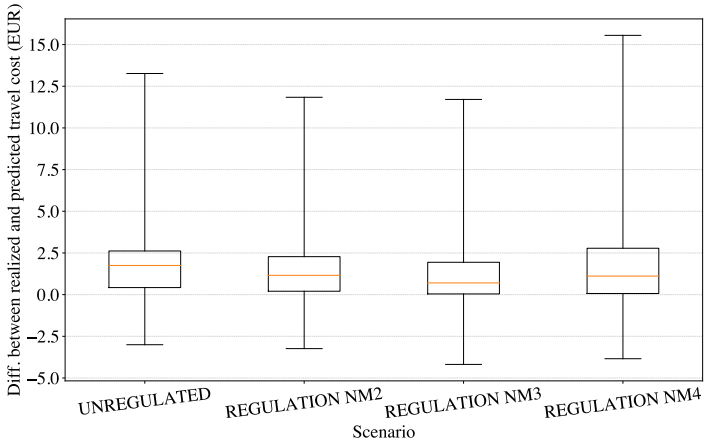


FIGURE 5.15 – *Statistics on differences between predicted and realized travel costs for the unregulated, regulation NM2, NM3, and NM4 scenarios.*

5.7.4 Fairness concerns

On top of reducing system-wide emissions and keeping a revenue-neutral pricing scheme, fairness is a natural objective for the regulator. It is not easy to establish a universal standard for fairness. Three questions should be answered to define the fairness principle to apply in a certain context according to Meyer & Roser (2009).

First, what kind of impact is considered? Here, we choose to look at the impact of the regulation on the travel costs of travelers. Second, how to categorize the population into groups? Defining social groups is not particularly relevant in our theoretical urban area since it would add more hypotheses. Spatial hypotheses have already been made to design the urban scenario. Hence, we choose a spatial criterion to group travelers. They are gathered per OzDz pair. Third, which principle rules to determine whether a distribution is fair? Li (2018) describes several fairness principles. Horizontal and vertical fairness respectively refer to the fair distribution of the regulation impacts within a group of travelers having the

same need or ability and across different groups. Under one of these fairness dimensions, they define opportunity and proportional equity. Opportunity equity requires all users to experience the same travel costs regardless of their personal conditions. Proportional equity consists in allocating the regulation impacts proportionally to the characteristics of travelers.

This section investigates the impacts of the regulation regarding two different definitions of fairness. The first one is based on the vertical and proportional principles: the average travel cost of each OzDz pair in the regulated situation should be impacted proportionally to the average travel cost of this OzDz pair in the no AMoD or unregulated situation. The standard deviation of the ratios between the average travel cost under regulation and the average travel cost under no AMoD or no regulation for all OzDz pairs can measure the fairness in the change introduced by the deployment of AMoD and the pricing scheme. A lower standard deviation means that different OzDz pairs experience closer effects on their average travel cost. This definition assesses the acceptability of AMoD and the policy by users. However, it is built on a comparison with the no AMoD or unregulated scenarios which may be unfair.

To capture the intrinsic fairness of the system, the second definition is based on the vertical and opportunity principles: the average travel cost for all OzDz pairs should be as close as possible. A lower Gini coefficient of average travel costs of OzDz pairs means a more fair distribution of travel costs among the groups of users.

5.7.4.1 Whole demand scenario

Table 5.5 summarizes equity metrics and associated indicators for the four scenarios analyzed in section 5.7.2.2. Compared with the no AMoD scenario, the deployment of AVs leads to slight increases (no more than 0.8%) of the average travel cost for 5% of the OzDz pairs, and to reductions up to 20% of average travel cost for 95% of the OzDz pairs. The mean of the ratios is smaller than 1, indicating a positive impact of the deployment of AVs, and their standard deviation remains small (3.88%).

Regulation WD1 introduces more disparities in the change of travel costs. The majority of OzDz pairs (59%) undergo an increase in their average travel cost, and the maximum increase reaches 34.8%. The standard deviation attains 10.7%. The indicators worsen for regulation WD2, with 68% of OzDz experiencing an increase up to 54.3% in their average travel cost. Comparing regulated and unregulated scenarios leads to similar results. The decrease in average travel cost concerns only 17% of OzDzs for regulation WD1, and 11% for regulation WD2. The 11% OzDzs all have their origin in $\{z_{0-1}, z_{1-0}, z_{1-2}, z_{2-1}\}$, namely in a central suburban zone gathering train, metro and bus lines terminal stations (see Figure 5.3). They correspond to traversing trips, i.e., with a destination in suburban zones located in the opposite direction.

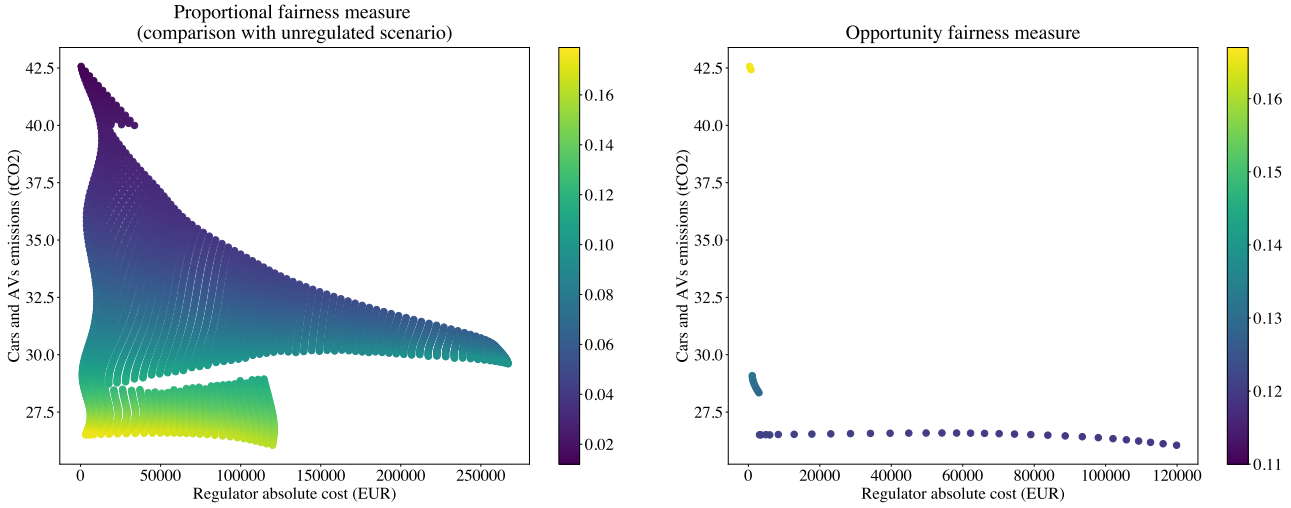
Given these results, both regulations are difficult to justify under the proportional fairness definition and may raise acceptability issues. However, they improve the intrinsic fairness of the system by reducing the spatial disparities in travel costs as shown by the reduction of the Gini coefficient.

We build a new surrogate model per fairness indicator based on the 250 simulations

launched for the multi-objective optimization of section 5.7.2.2. Figure 5.16 shows the estimated PF for the three objectives of the regulator with proportional and opportunity equity metric. It is clear from Figure 5.16a that emissions reduction conflicts with proportional fairness. Moreover, releasing the revenue-neutral objective does not help much in minimizing the proportional equity metric. On the contrary, emissions and opportunity fairness metric reduce in accordance (Figure 5.16b).

	No AMoD	**No regulation**	**Regulation WD1**	**Regulation WD2**
Proportional fairness: comparison with no AMoD scenario				
Min of ratios	-	0.808	0.798	0.769
Max of ratios	-	1.008	1.348	1.543
Percentage of ratios < 1	-	95%	41%	32%
Mean of ratios	-	0.925	1.017	1.079
Std of ratios	-	0.0388	0.107	0.161
Proportional fairness: comparison with the unregulated scenario				
Min of ratios	-	-	0.890	0.858
Max of ratios	-	-	1.392	1.651
Percentage of ratios < 1	-	-	17%	11%
Mean of ratios	-	-	1.101	1.168
Std of ratios	-	-	0.114	0.178
Opportunity fairness				
Gini coeff.	0.165	0.165	0.131	0.121

TABLE 5.5 – *Fairness of the regulations found for the whole demand scenario.*



(a) *Considering proportional fairness principle.*

(b) *Considering opportunity fairness principle.*

FIGURE 5.16 – *Estimated PF for the three objectives of the regulator on the whole demand scenario.*

5.7.4.2 Non-motorized travelers scenario: an origin-based pricing scheme

Table 5.6 provides the fairness indicators for the regulation highlighted in section 5.7.2.1. The same tendency emerges: the regulation decreases proportional fairness but increases opportunity fairness.

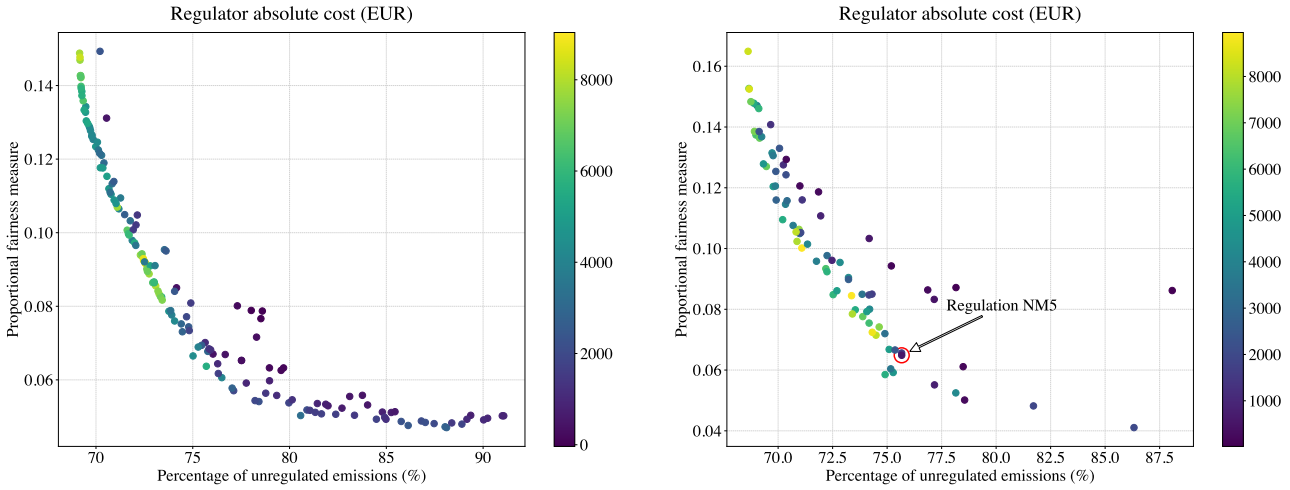
For this scenario, we relaunch the MOBOpt algorithm considering system-wide emissions, pricing scheme absolute cost, and proportional equity as objectives. Instead of considering a global tax on door-to-door AV rides and a global subsidy for intermodal AV rides, we make these prices dependent on the users' origin zones. Given the symmetry of the network and the demand scenario, zones can be grouped into three clusters: $\{z_{1-1}\}$ corresponding to the center, $\{z_{0-1}, z_{1-2}, z_{2-1}, z_{1-0}\}$ corresponding to central suburban zones, and $\{z_{0-0}, z_{0-2}, z_{2-2}, z_{2-0}\}$ corresponding to corner suburban zones. The same price values are applied in all zones belonging to the same cluster. $|\mathscr{P}|$ equals 6 for this origin-based pricing scheme. As in section 5.7.3, a penalty is added to the estimation of objectives leading to a higher carbon footprint compared to the unregulated scenario.

	No regulation	**Regulation NM1**	**Regulation NM5**
Proportional fairness: comparison with unregulated scenario			
Min of ratios	-	0.835	0.816
Max of ratios	-	1.448	1.191
Percentage of ratios < 1	-	44%	43%
Mean of ratios	-	1.047	1.013
Std of ratios	-	0.144	0.065
Opportunity fairness			
Gini coeff.	0.148	0.095	0.125

TABLE 5.6 – *Fairness of the regulation found on the non-motorized scenario.*

Figure 5.17 shows the estimated and observed PFs obtained with 400 simulations. The estimated PF profile suggests that we can find designs for the origin-based pricing scheme leading to significant carbon emissions reductions with a low impact on the proportional equity metric. It indicates a threshold in carbon emissions reduction (between 75% and 85% of the unregulated scenario emissions) above which the cost in proportional inequity becomes higher. Below this threshold, the revenue-neutral constraint is reasonably verified.

Figure 5.17b reveals a similar PF shape. We highlight one regulation among the 83 found by the algorithm, located within the relevant range discussed above. Regulation NM5 leads to 75.7% of the non-regulated scenario emissions, a proportional equity indicator of 0.065 (less than a half of Regulation NM1 score), and a pricing scheme cost of 584.2 EUR. Regulations NM1 and NM5 lead to the same proportion of OzDz pairs experiencing a lower travel cost than in the unregulated scenario. However, with the origin-based pricing scheme, the best improvement in average travel cost is slightly better, and the worse deterioration is slightly smaller. Notably, trips within the city centers undergo an 18% increase in average travel cost. Regulation NM5 also has a better opportunity fairness metric than the unregulated scenario.



(a) *Estimated PF.* (b) *Observed PF.*

FIGURE 5.17 – *Estimated and observed PF for the three objectives of the regulator on the non-motorized demand scenario.*

The corresponding prices are -2.07 EUR, -0.51 EUR, -0.53 EUR for p_{TAP}, and 3.92 EUR, 1.43 EUR, 1.12 EUR for p_{TA} in the center, central suburban, and corner suburban zones respectively. The regulation in the city center is stronger with higher prices values. Corner suburban zones benefit from a slightly higher subsidy and a smaller tax than central suburban zones.

Table 5.7 presents the changes in distances traveled per mode and can be compared to Table 5.3. Under regulation NM5, the market share loss of the train is reduced, and the total walking distance increased. It suggests that the pricing scheme could target the door-to-door AV riders more sharply and limit the mode shift from **P** to **AP**. Moreover, the ratio between EVKT and the total distance traveled by AVs passes from 19% under regulation NM1 to 17%.

Walk	+3.37%
Train	-5.26%
Metro	+36.5%
Bus	+59.0%
AV	-28.8%
Empty AV	+2.19%

TABLE 5.7 – *Changes in the total distances traveled per mode between unregulated and regulation NM5 scenarios for the case of non-motorized demand.*

5.8 Conclusion and discussion

In this chapter, we have approached the pricing scheme design problem with simulation-multi-objective optimization. We have described our simulation model, which is based on

the MnMS simulator but includes several contributions, such as: (i) the application of generalized cost functions taking into account regulation prices and the management of flexible intermodality in the mode and route choice model, (ii) the development of an AMoD mobility service where requests and vehicles are optimally batch-matched over a horizon. If the simulation approach allows to model in detail the interactions between travelers and AVs, and accounts for the short-term operational problems and constraints inherent to on-demand systems, it is a black box that requires significant time to run. Designing an optimized pricing scheme relevant to several objective functions obtained with simulation is challenging. We have applied a revised version of a state-of-the-art MOBO algorithm based on objectives meta-modeling with Gaussian Processes to deal with it. Our contribution lies in a knowledgeable exploration strategy using Gaussian processes' ability to estimate the prediction error made at any point of the search space. The resulting simulation-multi-objective optimization approach can handle complex pricing schemes and objectives.

The numerical experiments were conducted on an advanced theoretical multimodal network representing a large urban area, with up to six decision variables and three objectives. Five price types gathered into three regulation policies and five objective functions gathered into three multi-objective problems have been investigated. Applying the proposed methodology and analyzing the resulting observed PF allowed us to draw four types of conclusions regarding the regulation of AMoD.

First, in a system where demand is flexible, i.e., travelers have access to various modes of transportation, including the personal car, the regulation of AMoD should be studied by the regulator jointly with the regulation of the most polluting modes or counterproductive mode shifts may emerge. We found that in our study case, cars should be taxed at least as much as door-to-door AV rides for the pricing scheme to generate benefits in terms of carbon footprint.

Second, when the pricing scheme includes taxes on door-to-door AV rides and subsidies for encouraging synergy between PT and AMoD, we can find a quasi revenue-neutral design that generates nearly all achievable benefits regarding system-wide carbon emissions. When the car's captive share of the demand is removed, subsidizing more than taxing generates negligible carbon emissions savings, and taxing more than subsidizing is not relevant (Pareto optimal). When the whole demand is considered, taxing more than subsidizing achieves negligible carbon emissions savings, and subsidizing more than taxing is not relevant. Consequently, the arbitration between the different policies composing the PF is easy. The selected regulations largely exploit the synergy between PT and AMoD via intermodality. AVs feed the PT network such that the collective and active distance-based mode share under regulation is higher than the baseline scenario without AMoD. Significant carbon emissions savings are achieved: 30% on the non-motorized demand scenario and up to 39% on the whole demand scenario. However, the selected regulations disadvantage the AMoD operator and the travelers.

Third, the AMoD operator profit decrease comes from modifying the demand pattern for AMoD: the AVs fleet is less efficient in serving door-to-station/station-to-door requests than door-to-door requests. In our theoretical study case, the ratio between empty VKT and total distance traveled by AVs passes from 12% without regulation to 19% under regulation.

Fourth, to reduce carbon emissions with a quasi revenue-neutral pricing scheme, sep-

arating for-travelers and for-AVs prices helps to limit AMoD operator profit decrease and total travel cost increase. On the non-motorized demand scenario, we could find a policy reducing emissions by 28% with no more than a 5.5% reduction in the AMoD operator profit and increase in the total travel cost.

Five, the Pareto optimal regulations found contribute to improving the spatial opportunity equity in the system. However, they may raise acceptability issues since some OzDz pairs experience a great increase in their average travel cost. In contrast, others benefit from a large decrease. In our case study, the favored OzDz pairs correspond to long intra-suburban trips. Differentiating price values depending on users origins helps limiting the proportional inequity.

This work could be extended to investigate many other regulation policies and objectives. In terms of regulation, the simulation model allows the definition of complex pricing schemes, but the number of decision variables is limited to 20, which is the known limit for BO. To overcome this limitation, one can use parameterized pricing schemes as it is done in Dandl *et al.* (2021) or Shou & Di (2020).

A direct extension of this work is the application of the framework to real urban areas. The network of Lyon city is envisaged to confirm the tendencies highlighted in the theoretical study case and evaluate the potential of intermodal AMoD.

MOBOpt could be improved in several ways. Notably, an adaptative scheme could be defined to adapt the algorithm parameters (π, w) to balance three objectives along the optimization procedure: (i) find Pareto optimal points, (ii) explore under-explored regions of Ω, (iii) densify the observed PF. Such an adaptative scheme may help the algorithm to converge quicker to a good quality dense PF.

Regarding traffic dynamics, even if early studies assumed that the MFD is a feature of the network only, the OD matrix and user route choices affect its form in practice. In the scenario focused on non-motorized travelers, the background traffic is given and assuming a fixed MFD form is a reasonable hypothesis. However, the regulation policy may substantially modify the demand pattern for the whole demand scenario. Encouraging to have intermodal trips (**AP** or **CP**) may lead to heavy heterogeneity in terms of congestion distribution (mainly when a few PT hubs attract most of the demand) and affect the MFD definition (Buisson & Ladier, 2009). Optimizing the regulation policy and adapting the MFD relationship jointly is challenging. Estimating the MFD form under a different demand pattern could be achieved through micro-simulation. Such an estimation is time-consuming. Hence, limiting the number of re-estimations would require detecting when the modification in the demand pattern is substantial enough to provoke a major change in the traffic dynamics.

Conclusion of Part III

This part addressed the pricing scheme design as a black-box multi-objective optimization problem. We focused on setting up a relevant simulation model and coupling it with an efficient optimization algorithm. Significantly, the simulation model is predictive, with a horizon for mode-route choice and AVs dispatching, and constrained, with time windows preventing a major deviation from the system equilibrium.

The methodology allowed us to highlight the relations between several objectives related to system efficiency, sustainability, and equity in a theoretical city. If some results specifically apply to the network and demand pattern considered, others are stand-alone. For example, the on-demand fleet efficiency reduction as it answers more intermodal requests is intuitive and in line with some analytical studies (Bimpikis *et al.*, 2019). The proposed framework acknowledges this efficiency loss, CO2 emissions reduction, and other implications of the regulation for users and AMoD operators numerically.

There are still gaps to be filled for future exploitation. As for Part II, the regulation design has been done under one demand scenario, representing the typical demand during peak hours. Out of these nominal conditions, the price values found may need a revision. One should relaunch the optimization procedure to adapt to other typical demand scenarios. One option is building an exhaustive catalog of nominal demand profiles with the corresponding optimal price values. It implies having access to daily demand data, extracting recurrences in the patterns, and launching many optimizations.

Such a methodology does not accommodate extreme situations as policies are pre-built. Another option is a dynamic pricing scheme integrating a feedback mechanism to detect and adapt prices to changes in the demand profile. It is very challenging and calls for different techniques than the ones used in this part, though.

The work presented in this part remains exploratory and deserves extensions for operational exploitation.

Part IV

Conclusion

6. Global overview, findings and future research

6.1 Summary and global overview

Motivated by shared mobility's interest in making the urban transportation system greener, this thesis focused on a specific declination of shared mobility with AVs called AMoD. The environmental benefits of AMoD deployment depend on the integration of AMoD and PT. This thesis aimed to study different regulation measures to encourage synergy between PT and AMoD and exploit the full potential of intermodal AMoD. Three goals were pursued in the manuscript: (i) account for and quantify the benefits of intermodal AMoD regarding system efficiency, sustainability, and equity indicators, (ii) describe the circumstances of cooperation or competition between PT and AMoD, (iii) optimize the studied regulation measures.

The regulation of MoD systems has received much attention in the literature, but the study of AMoD-specific regulation is relatively new and rarely approached with optimization or multi-objective optimization. Moreover, the few studies dealing with AMoD regulation optimization rather exploit the possibility of sharing AV rides rather than the possibility of transferring between PT and AMoD. This thesis examined AMoD-specific regulation optimization in light of the intermodality between PT and AMoD, which is an original line of research.

As commuting journeys represent a substantial part of the urban demand, challenge the personal car model because of congestion, and are recurrent, we chose to restrict to regulation policies built on the morning peak. Parts II and III of the thesis introduced two study cases for which we consider distinct regulation types. The corridor study case involves two heavy linear transportation infrastructures attracting travelers and vehicles flows. In this case, we optimized a regulation impacting the design and operation of AMoD jointly with the MRT line design. It consists in segmenting the AMoD coverage zone along the corridor. On the contrary, the large urban area study case is characterized by meshed PT and road networks with many possible transfer points. In this case, we optimized multiple sets of regulation prices addressed to travelers riding AVs and to AVs themselves.

Three steps were required to achieve the goals of the thesis: modeling, simulation, and optimization. In terms of modeling, the distinct nature of our study cases led us to make different hypotheses for them. Figure 6.1 highlights these differences.

	CORRIDOR	GREATER URBAN AREA
NETWORK	Simple, linear	Generic, meshed
INTERMODALITY	Limited	Flexible
NETWORK LOADING DYNAMICS	Detailed description of the loading phases	Macroscopic loading model
AMOD MODEL	Model-based approximation	AV-centric operational policy, detailed analysis
SUPPLY-DEMAND COUPLING	Dynamic User Equilibrium resolution method	Instantaneous mode and route decision based on current system state
OPTIMIZATION	Single objective and consecutive optimizations, genetic algorithm	Multi-objective Bayesian Optimization

FIGURE 6.1 – *Comparison of the approaches chosen for the corridor and the urban area case studies. Green indicates the main contributions in each of the cases. Orange indicates the compromises made in this thesis and would deserve to be reworked in further research.*

Modeling In the corridor study case, the network is quite simple and characterized by a linear configuration causing the emergence of substantial delays. Consequently, the network loading dynamics and the coupling between supply and demand were carefully addressed with a simple model incorporating time-dependent mode and route choice subject to user equilibrium constraints. Three variants of the model have been introduced. The linear, pseudo-dynamic, and dynamic variants cover an increasing number of aspects of the system: the morning commute dynamics, travelers disaggregation, and endogenous time-variant AMoD service time. The linear corridor model enabled us to precisely describe how UE settles and evolves as the network loads and to explicit the synchronization and desynchronization effects between bottlenecks. It helped us understand the circumstances of cooperation and competition between PT and AMoD and the effect of segmenting the fleet coverage zone.

In the dynamic model, the reduced number of drop-off spots and the many-to-one demand pattern hinder the AMoD operational leeway to optimize AVs dispatching to maximize profit. Hence, a model-based approximation of the AMoD service time rather than an AV-centric AMoD model was used. If this assumption led to a consistent service time profile following the loading of the system and allowed to reach a good quality UE in a few iterations numerically, it would deserve to be validated with a detailed post-analysis of AMoD operation for serving the identified demand

In the urban area study case, the spatial extent of the network, its geometry, and the urban demand pattern special features (imbalances) widen the operational leeway of AMoD.

It raised the need for a detailed AMoD model in an agent-based framework accounting for routing, matching, and relocation constraints. Hence, we introduced several variants of an AMoD operational policy called batch-matching over horizon to model the profit-oriented behavior of the fleet of AVs in an environment subject to for-AV taxes and subsidies. We analyzed their scope (i.e., the conditions under which the fleet is well managed) and their fallout for travelers regarding the quality of service and spatial equity. We showed that the policy performs better than standard event-based heuristics such as the *nearest-idle-vehicle* strategy.

Meanwhile, with the meshed network, the congestion is expected to be more evenly distributed across the links of the network, delays should be smaller and impact less travelers' mode and route choices. Hence, the network loading dynamics were addressed at a macroscopic scale. Moreover, instead of computing the equilibrium of the system, we considered that travelers choose their mode and route based on the instantaneous state of the system without adapting their choices regarding their experiences. This assumption led to differences between predicted and realized travel costs remaining below +2.78 EUR (resp. +2.34 EUR) for three-quarters of travelers in the non-motorized (resp. whole demand) scenario. The increase between predicted and realized travel costs is limited to +35% (resp. +22%) for three-quarters of travelers in the non-motorized (resp. whole demand) scenario. It indicates that the assumption deserves to be reworked while maintaining compatibility with the optimization techniques used.

Simulation We described the simulation platforms used in the manuscript. In chapter 4, we designed a simple simulator including components to model AMoD interactions with travelers within a multimodal transportation network. In chapter 5, we described how we adapted a comprehensive agent-based simulation platform (MnMS) to integrate generalized travel cost functions, regulations prices, the management of flexible intermodality, and the AMoD batch-matching over horizon operational policy.

Optimization In terms of optimization, two techniques were mobilized. The small computational burden of the DUE resolution method proposed to deal with the corridor study case allowed us to use an evolutionary approach. The genetic algorithm developed demonstrated good performance to optimize the joint AMoD-PT design. If several objectives for the regulator were studied across distinct single-objective optimization experiments, the diversity of stakeholders and their respective objectives were considered with a liberalism scenario.

The pricing scheme design was formulated as a multi-objective optimization problem in the urban area study case. As population-based algorithms were not adapted to our simulation-multi-objective optimization approach, we applied a recent methodology for efficient multi-objective optimization of time-expensive black-box functions. The methodology belongs to Multi-Objective Bayesian Optimization (MOBO). It relies on Gaussian Processes to meta-model the objective functions of the problem. It showed good performances in our numerical experiments by reaching multiple quality regulation policies in a relatively small number of simulations for problems with up to six decision variables and three objectives. While Bayesian Optimization is increasingly used in transportation studies, the appli-

cation of MOBO in simulation-based transportation problems is rare (Wang *et al.*, 2023). Our work contributed to develop this line of research.

6.2 Major operational outcomes

The work presented in this thesis answered the three goals initially announced and led to several operational outcomes regarding the regulation of AMoD. The major conclusions for the corridor case are listed below.

- When the corridor is studied with a many-to-one demand pattern, the DUE becomes more complex than with a one-to-one demand pattern. Compared to Laval (2009), desynchronization effects between bottlenecks occur. Upstream bottlenecks get out of sync due to the overloading of downstream bottlenecks. These effects constitute an opportunity to exploit in fostering synergies between PT and AMoD.
- The corridor design parameters, especially the number of collectors and their locations, largely influence the dynamics of the system. When a few collectors are well spread along the corridor, the direct first-mile pattern is favored, and synergy between PT and AMoD emerges. When collectors are numerous, or their spacing decreases downstream, the long first-mile pattern intensifies and the PT line loses market share.
- For each of the three objectives considered, namely maximize the MRT use, minimize the average travel time, and minimize car use, the joint optimization of the corridor design and the AMoD regulation led to better results than the protectionism and liberalism scenarios, which justifies the need for regulation. However, the regulation implies less uniformity in the delays experienced by travelers along the corridor due to congestion and AMoD service time.
- The one fleet per collector design strategy for AMoD is highly relevant for increasing suburban train lines' attractiveness.

The major outcomes in the urban area are listed below.

- The batch-matching over horizon operational policy allows better optimization of the fleet than using event-based rules, but the horizon management, utility definition, and horizon length are crucial parameters to maximize the AMoD operator profit.
- Encouraging intermodal AV rides and discouraging door-to-door AV rides leads to tougher imbalances in the AMoD demand pattern, which requires higher relocation efforts for the fleet. More kilometers are traveled empty for fewer kilometers served, but the AMoD operator profit does not necessarily falls because more requests can be served.
- We can find quasi revenue-neutral pricing schemes based on a tax for door-to-door AV rides and a subsidy for intermodal AV rides that generate nearly all achievable benefits regarding system-wide carbon emissions. Savings of 30% and 39% are reached when regulation applies to non-motorized and all travelers, respectively. Separating for-travelers and for-AVs prices helps to limit AMoD operator profit decrease and total travel cost increase. Differentiating prices by users' origins helps to limit proportional inequity compared to the unregulated scenario.

6.3 Research perspectives

The work could be extended in various manners. The following list summarizes the direct extensions possible to validate or precise our results.

- As stated in section 6.1, each study case has made a compromise in modeling. The first direct extensions consist in overcoming these limitations. The model-based approximation of the AMoD service time used in the corridor could be validated by comparing the service time profile obtained at the end of the MSA procedure with the service time profile obtained by building the routes of AVs. The built routes should address the demand for AMoD within the traffic conditions evaluated by our dynamic model. Regarding the urban area, we have restricted the computational complexity of the simulation by considering mode and route choice based on the instantaneous state of the system without computing a user equilibrium. Adding an iterative procedure to re-assign users depending on their previous traveling experiences, similar to the co-evolutionary principle used by MATSim, may lead to different results and help refine the evaluation of carbon emissions reduction.

- In Chapter 1, we worked with the assumption that travel times mainly rule commuters' mode and route choices. In Chapter 2, we included the monetary aspect by introducing fares, operational costs of the different modes, and values of time per type of activity. Values of time were taken from Kolarova *et al.* (2018) stated choice experiment realized on a representative sample of the German population. A deterministic choice model was applied in both chapters. Moving a step forward in modeling commuters' behaviors (their preferences and willingness to use AMoD) implies (i) using a more advanced choice model, such as logit or nested logit one, which better renders the stochasticity of travelers' choices, and (ii) calibrating utility functions on the data of the studied city. Applying our framework to an actual city would require an additional stated preference survey to capture behavioral factors specific to the geographical region in question, as it is done in Scheltes & de Almeida Correia (2017), Yap *et al.* (2016). Indeed, using results from the existing literature is not reliable since they largely depend on the context. Despite their ability to capture some behavioral factors, such experiments can only capture an a priori acceptability level of AMoD. Field experiments are the only way to get a more accurate idea of the acceptability in use. Still, they can help designing more robust regulation policies. Coupling a more advanced choice model with our framework is challenging. Indeed, if MOBO with Gaussian Processes and the proper kernels can deal with noisy functions, when noise is substantial, the approach may fail to sample non-dominated points. More advanced infill criteria than the one presented in this work should be considered to account for noise and better guide the sampling (Daulton *et al.*, 2021).

- As mentioned in the conclusion of chapter 5, the simulation-multi-objective optimization approach proposed can be applied to a real network. To draw reliable conclusions from such a numerical application, calibrating the mode choice and the traffic models is a required preliminary step.

- As we focused on ride-hailing with AVs and PT-AMoD cooperation to derive environmental benefits for the system, a direct extension to generate more benefits is to

include the possibility of sharing AV rides.

- If each regulation type studied befits the case in which it is considered, both could be mixed in large urban areas. The agent-based simulation platform could help evaluate a hybrid regulation based on a pricing scheme and AMoD geofencing around train stations in the suburbs.

Additionally, the work opened the way to new research questions.

- This thesis focused on the morning commute but the economic viability and operational policy of AMoD could be assessed over longer periods. The demand levels across a typical week generally influence the fleet size and fare choice. The management of a fleet of electric AVs, and more specifically, the charging policy, is to be considered on a daily basis rather than on isolated peak events. Extending the studied period may influence the AMoD behavior and, consequently, the results obtained in part III. It may impact the relevancy of the studied regulation policies and bring to the fore the question of temporal synergy between PT and AMoD (synchronization of PT and AMoD supply in time) on top of the spatial synergy we discussed in this thesis. The objectives conflicting in the short term may not be in conflict in the long run.

- In the framework proposed in chapter 5, the AMoD operator only has short-term operational decision variables to maximize its profit. In future research, we will consider adding strategic decision variables for the AMoD operator (e.g., fleet size, fares, horizon length) and compare different "playing modes" in the game bringing the regulator's environmental considerations and the AMoD operator profit into conflict. We identify three interesting modes:

  - cooperation, in which the players seek an agreement on the Pareto Front of their respective objectives. They act so the other's objective cannot be improved without degrading their own objective. The methodology presented in chapter 5 applies in this case.
  - hierarchical competition, in which one player leads while the other follows. The leader evaluates how the follower will answer once it has observed the leader's move. The leader chooses the action that maximizes its payoff, and the follower behaves as predicted. It corresponds to a Stackelberg equilibrium and can be addressed with a bi-level optimization framework, which is compatible with Bayesian Optimization. The hierarchy in the competition somehow represents the difference in adaptability between the players. The follower adapts quickly to a certain situation, and the leader has a longer-term vision. The standard top-down approach places the regulator as the leader. In contrast, a bottom-up approach would place the AMoD operator as the leader.
  - standard competition, where the players act simultaneously and selfishly with a correct estimation of the other's choice. Computing Nash equilibria in the context of time-expensive black-box functions is challenging. However, recent works on this subject could help in this task Picheny *et al.* (2019).

  Comparing these models leading to a priori different equilibria would be insightful. Similar investigations could be realized in the corridor study case, where both stakeholders already have identified decision variables.

- Another element we have yet to approach is the dynamics of the regulation and economic strategies of stakeholders. Considering dynamic decision variables rather than static ones and long-term objectives rather than short-term ones would allow for capturing more realistic strategies. One relevant example is the temporary reduction of AMoD fares at a loss to gain market share durably and increase fares in the future. Many scientific challenges accompany this question: including a life-cycle assessment in the stakeholders' objectives and travelers' behavior modeling in such a varying context.

- In the introduction, we presented the main differences between MoD and AMoD. We claimed that MoD-specific regulations are not necessarily adapted or efficient in the context of AMoD because of these differences. We can reverse the question: are the AMoD-specific regulations adapted for the MoD context? Some conceptual and methodological choices made throughout the thesis are questionable when on-demand vehicles are human-driven. First, the fixed fleet size assumption ignores the entry and exit of drivers onto the market. The variation in supply is essential and should be taken into account. Second, the model-based approximation used in Chapter 2 to evaluate the effective service time profile implies that vehicles comply with the order in which requests appear, even if a long empty repositioning is needed. Similarly, the dispatching strategy formulated in Chapter 4 could be more robust to stochasticity in vehicle behavior. In Chapter 5, the simulation model entails only two types of agents. In the context of MoD, it should include drivers. Moreover, when it comes to balancing the reduction in CO2 emissions, the loss of profit for AMoD, and the increase of travel costs for users, the profit aspect should be split into AMoD operator profit and drivers welfare. The question of equity underlined for commuters is also valid for drivers. For these reasons, the results have no direct applicability to MoD systems. The frameworks would require adjustments to render the new complexity introduced by the consideration of drivers decision-makers. Also, from the standpoint of traffic engineering, we could go further into modeling autonomous driving impacts on the system.

A. APPENDICES FOR CHAPTER 3

A.1 Sensitivity of the results on the other parameters of our dynamic model

Among the different parameters of the dynamic model, only four served as decision variables for optimization in Chapter 3. Freeway and street speeds, MRT cruising speed, headway and dwell time, bottleneck capacities, and AVs ratio are considered fixed but do impact the results. The complete sensitivity analysis of these parameters on outputs could help validate certain conclusions drawn from the West Lyon corridor. We did not proceed a comprehensive analysis, but the knowledge of the principles ruling the system's dynamics allows us to formulate assumptions on the impact of these parameters.

Freeway and street speeds, MRT cruising speed, headway, and dwell time affect the free-flow travel times of the different alternatives. Hence, they may change the order in which options start to be used, and they change the gaps in travel time between two consecutive options. If gaps only are affected, the values of the final distance-based mode shares should be modified, but the nature of cooperation-competition regimes for a particular design is maintained. If options order changes, the nature of cooperation-competition regimes emerging from a specific design may differ.

Bottleneck capacities do not affect the order in which alternatives are activated but the diversion pattern type, the number of diversion itineraries mobilized over time, and the final amount of users diverting. The regulator cannot easily modify the capacity of the downstream collector, but the capacities of AV-to-MRT transfer points would make an interesting lever for the regulator.

Similarly, AVs ratio was considered the same in all coverage zones in Chapter 3. AVs ratio represents the number of AVs operating in a fleet on the number of travelers living in its coverage zone. As the AVs ratio decreases, AMoD supply becomes insufficient, and on-off effects in requests and oscillations in service time appear as shown in figure A.1). As soon as **a** extra travel times overcome w_0 due to a substantial T_s, the service is deserted for a moment, then used again. A smaller AVs ratio results in fast variations of $\widehat{T_s}$. These $\widehat{T_s}$ peaks are more difficult to reproduce by MSA, and a higher number of iterations is required to solve the fixed point problem on T_s. Reducing the AVs ratio has similar effects as geofencing but less strictly and makes the convergence more difficult. Consequently, it is not chosen

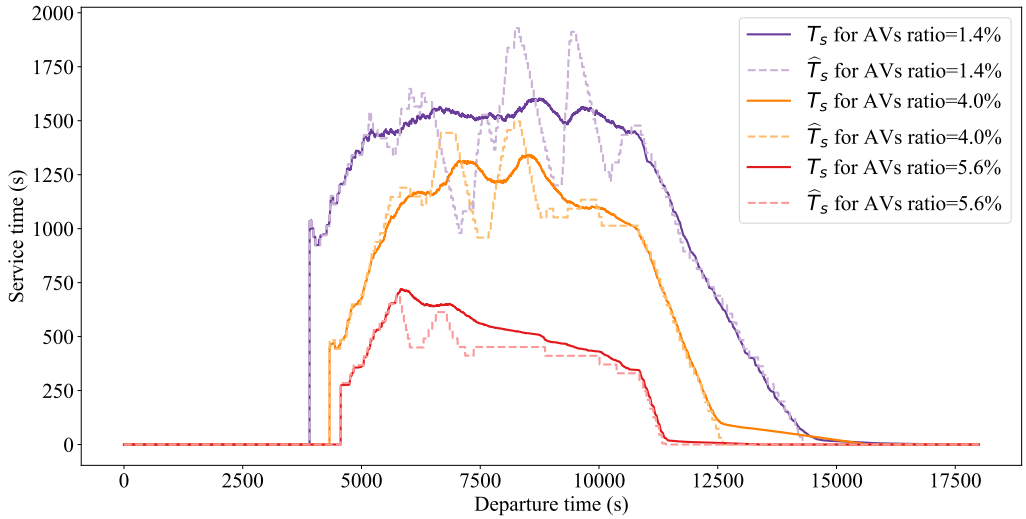


FIGURE A.1 – *Sensitivity of service time profile on AVs ratio, oscillations observed for low ratios are due to on-off effects.*

here as a decision variable, but a sufficient AVs ratio is chosen to prevent on-off effects.

The demand profile also influences the system's dynamics via bottlenecks desynchronization. Additionally to the West Lyon case, we applied the framework to three theoretical corridors with different distributions of commuters' origins. The homogeneous, sprawled, and commuter corridor types have homogeneously distributed origins, normally distributed origins with more departures from the downstream, and normally distributed origins with more departures from the upstream, respectively. Tables A.1, A.2, A.3, present the numerical results obtained for these corridors.

One can identify common points in the four scenarios. Opportunism always performs better than the other two policies on the indicator associated with the regulator's priority objective. AMoD designs emerging from opportunism to maximize the MRT usage favor more numerous fleets with smaller coverage zones. Geofencing for theoretical corridors is less strict than for the realistic one: some fleets extend on several vehicle attraction areas. Liberalism always performs better in terms of delays' uniformity.

Scenario	**Homogeneous city**							
Priority objective	Max. MRT usage			Min. avg. travel time			Min. car usage	
Policy	Protectionism	Opportunism	Liberalism	Protectionism	Opportunism	Liberalism	Opportunism	Liberalism
M	11	10	9	18	28	17	8	15
Number of fleets	0	6	1	0	12	1	2	2
c mode share	38.2	36.0	35.6	37.2	35.4	35.4	35.6	36.9
r mode share	61.8	16.7	28.9	62.8	0.4	38.5	22.2	27.8
a mode share	0.0	47.3	35.6	0.0	64.2	26.1	42.3	35.3
Car dist. mode share	27.8	22.3	23.9	41.4	45.4	34.7	22.4	24.1
MRT dist. mode share	64.5	69.4	61.9	52.7	46.8	40.3	67.3	57.9
AV dist. mode share	0.0	6.4	11.5	0.0	7.8	21.9	7.9	15.3
Walk dist. mode share	7.7	1.8	2.6	5.9	0.0	3.1	2.3	2.7
Avg. TT per trav.	35:06	28:36	30:22	30:18	20:50	27:13	29:43	33:54
Gini coef.	0.78	0.66	0.59	0.77	0.71	0.66	0.61	0.59
AV occupancy	-	7.2	3.6	-	6.6	2.6	6.1	3.5

TABLE A.1 – *Homogeneous city results.*

Scenario	**Commuter city**							
Priority objective	Max. MRT usage			Min. avg. travel time			Min. car usage	
Policy	Protectionism	Opportunism	Liberalism	Protectionism	Opportunism	Liberalism	Opportunism	Liberalism
M	14	7	9	16	21	14	9	9
Number of fleets	0	5	1	0	7	3	4	2
c mode share	37.7	36.7	36.6	37.9	35.4	35.9	35.6	36.4
r mode share	62.3	25.6	26.8	62.1	2.5	23.1	19.9	16.0
a mode share	0.0	37.8	36.7	0.0	62.1	41.1	44.5	47.6
Car dist. mode share	29.3	26.3	30.7	38.5	40.6	30.6	26.6	27.6
MRT dist. mode share	66.3	67.9	62	57.8	53.8	59.5	68.0	59.5
AV dist. mode share	0.0	4.1	5.9	0.0	5.6	8.9	4.4	11.6
Walk dist. mode share	4.4	1.7	1.4	3.7	0.1	1.0	1.0	1.3
Avg. TT per trav.	35:43	33:35	30:15	31:26	24:04	27:54	29:19	34:18
Gini coef.	0.78	0.63	0.59	0.77	0.66	0.56	0.59	0.63
AV occupancy	-	5.3	3.7	-	6.2	4.1	5.8	4.8

TABLE A.2 – *Commuter city results.*

Scenario	**Sprawled city**							
Priority objective	Max. MRT usage			Min. avg. travel time			Min. car usage	
Policy	Protectionism	Opportunism	Liberalism	Protectionism	Opportunism	Liberalism	Opportunism	Liberalism
M	18	14	10	25	32	23	7	12
Number of fleets	0	12	1	0	11	3	4	2
c mode share	37.4	35.1	35.1	36.3	34.4	34.2	35.4	35.4
r mode share	62.6	18.2	27.6	63.7	1.1	32.6	31.1	32.1
a mode share	0.0	46.7	37.4	0.0	64.5	33.2	33.5	32.5
Car dist. mode share	27.7	23.4	23.9	45.0	48.0	34.7	22.7	23.5
MRT dist. mode share	63.1	68.2	61.6	48.6	41.8	34.5	65.5	56.5
AV dist. mode share	0.0	6.3	11.4	0.0	10.1	28.0	7.2	15.4
Walk dist. mode share	9.2	2.1	3.0	6.4	0.0	2.8	4.6	4.6
Avg. TT per trav.	29:48	21:31	25:05	24:13	17:01	20:39	26:12	28:47
Gini coef.	0.78	0.71	0.59	0.78	0.69	0.61	0.67	0.62
AV occupancy	-	7.0	3.7	-	6.6	3.3	5.6	3.2

TABLE A.3 – *Sprawled city results.*

B. APPENDICES FOR CHAPTER 4

B.1 Shortest path algorithm in the agent-based simulation platform of chapter 4

There are two situations in which we need to compute a shortest path on the multimodal graph.

When AMoD builds the service plans and repositioning moves, it has to compute the shortest path for an AV to join the destination node in the graph. In that way, the deficit associated with the total distance traveled by empty AVs is minimized. We use an A-star algorithm with the manhattan distance as heuristic to proceed.

When a traveler chooses its itinerary, it has to compute the path with the lowest generalized travel cost. A transfer penalty applies each time the traveler changes mode, given that walk is not considered as a mode. For example, if a traveler walks till a bus station, boards, rides, alights it, walks to a subway station, boards, rides and alights it, then, walks till its final destination, only one transfer penalty is counted to account for the discomfort of changing mode from bus to subway. The transfer penalties between G_s and G_b are directly carried by the *alightboard* arcs. The other transfers necessarily pass through G_w. The corresponding penalties cannot be directly carried by the transfer arcs since the application of the penalty depends on the path taken till then. A traveler passing through a *board* or a *pick-up* arc undergoes the penalty only if she has used one of AV, car, train, subway, bus modes earlier in her journey. A traveler boarding an AV or a public transportation rolling stock after having walked from her home will not undergo the penalty. Consequently, we cannot use the traditional A-star or Dijkstra algorithms.

We implemented an adapted version of the Dijkstra algorithm. The traditional version of Dijkstra algorithm would explore the graph by inspecting in priority the nodes labelled with the lowest generalized travel cost (cumulative weight from origin to this node on the shortest explored path). In the revised version, we add a boolean in the node label. It equals 1 if one mode among AV, car, train, subway, bus has been used on the shortest explored path till this node, and 0 otherwise. This boolean allows to take into account a transfer penalty when needed, i.e., when an arc *board* or *pick-up* is passed through and the boolean equals 1. The graph is explored by inspecting in priority the nodes labelled with: (i) the lowest generalized travel cost, (ii) the lowest boolean. If two paths with the same generalized travel

costs are found till a certain node, the path with the lowest boolean is kept. In that way, we delay the moment from which transfer penalty is active. The adapted version of the Dijkstra algorithm is detailed in 5.

Algorithm 5: Adapted version of the Dijkstra algorithm

Data: Origin node o, Destination node d, Multimodal graph $G = (N, A)$ with generalized costs as weights
Result: Shortest path, Total weight on this shortest path

Initialize the temporary cost label for o ($c[o]$) to 0;
Initialize the boolean label for o ($b[o]$) to 0;
Initialize the ordered list of nodes to visit P with o (P keeps its nodes n sorted by increasing ($c[n]$, $b[n]$);
while $P \neq \emptyset$ **do**
    Get v the first node in P and remove it from P;
    Set the permanent cost label of v ($l[v]$) to $c[v]$;
    if *v is d* **then**
        break
    for *node u in the successors of v in G* **do**
        Compute the boolean label of u passing through arc (v,u), b_u = min(1, $b[v]$ + $b(v,u)$) where $b(v,u)$ equals 1 if (v,u) is an *av*, *car*, *train*, *subway*, or *bus* mode and 0 otherwise;
        Compute the cost label of u passing through arc (v,u), $c_u = l[v]$ + weight(v,u) + $\gamma_{\text{transfer}} b_u p(v,u)$ where $p(v,u)$ equals 1 if (v,u) is a *board* or *pick-up* arc and 0 otherwise;
        if *$l[u]$ is defined and $c_u == l[u]$ and $b_u < b[u]$* **then**
            $b[u] \leftarrow b_u$;
            Set v as the predecessor of u ($p[u] \leftarrow v$);
        else
            if *$c[u]$ is not defined or $c_u < c[u]$* **then**
                $c[u] \leftarrow c_u$;
                $b[u] \leftarrow b_u$;
                Set v as the predecessor of u ($p[u] \leftarrow v$);
                Add u to P while keeping the list correctly ordered;
            else
                if *$c_u == c[u]$ and $b_u < b[u]$* **then**
                    $b[u] \leftarrow bool$;
                    Set v as the predecessor of u ($p[u] \leftarrow v$);
Build the shortest path from d to o with the predecessors p;
Return the shortest path and $l[d]$;

B.2 Traveler behavior flow chart

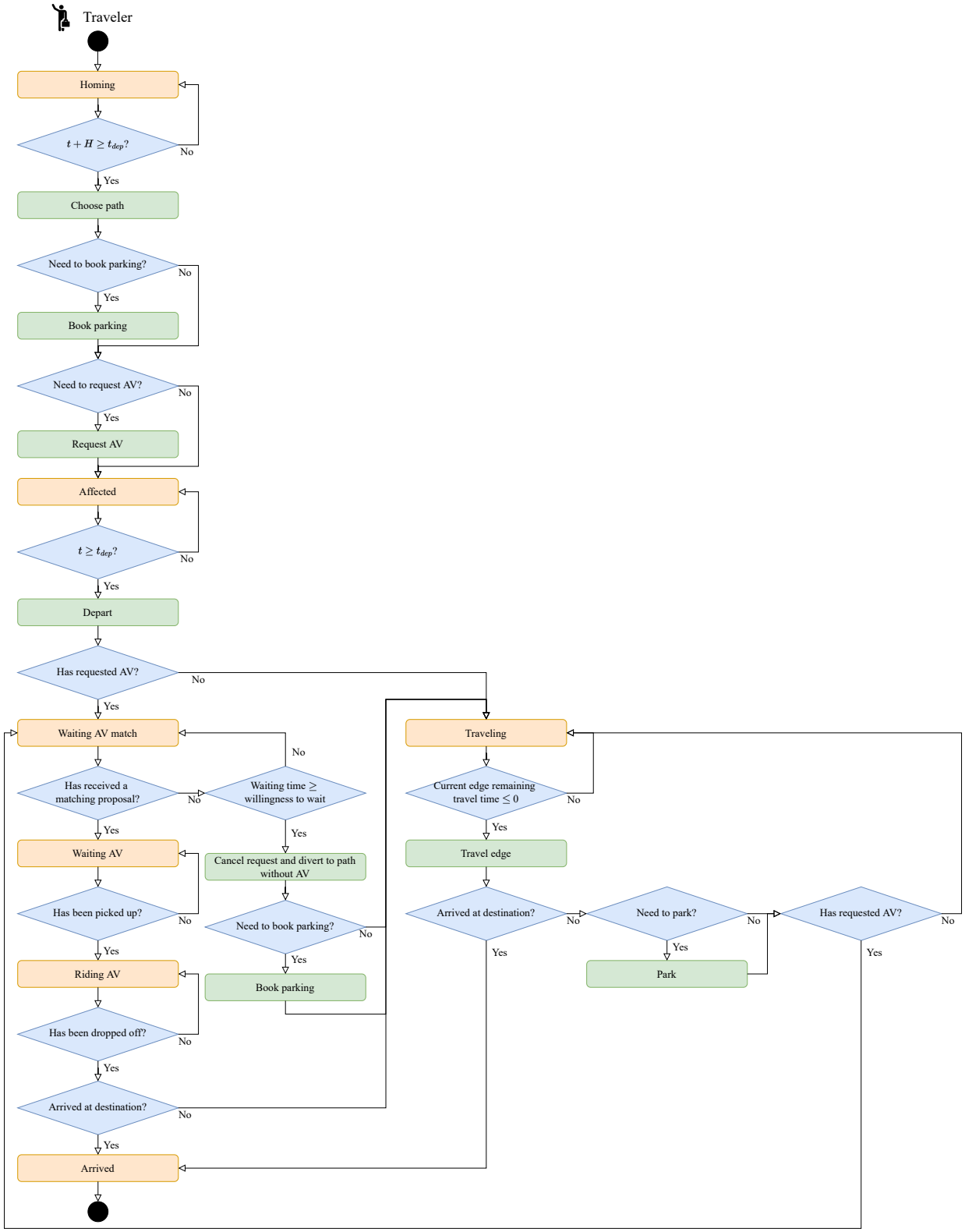


FIGURE B.1 – *Flow chart of traveler agent in the agent-based simulation platform used for chapter 4.*

B.3 AV behavior flow chart

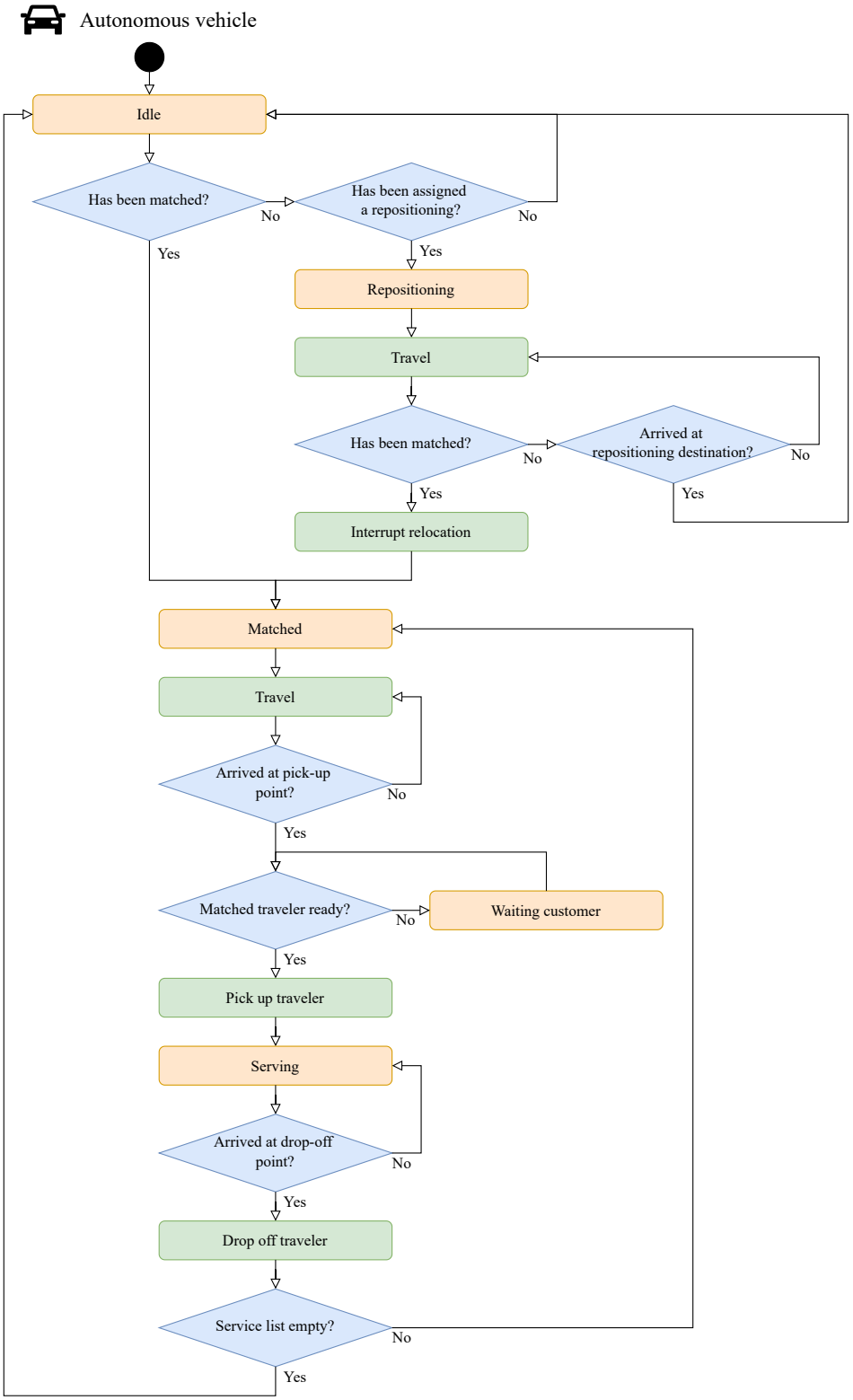


FIGURE B.2 – *Flow chart of AV agent in the agent-based simulation platform used for chapter 4.*

B.4 Batch matching in horizon with pre-built routes

In this appendix, we describe how the methodology proposed by Alonso-Mora *et al.* (2017a) can be applied to our case, i.e., with a profit-oriented operator and no ride sharing. Algorithm 6 is an adaptation of algorithm 2 with the preliminary building of routes to enable an AV to be matched with a group of request instead of one request at most. Steps 3, 4 and 5 are detailed below.

Note that K can be calibrated to reduce the computation time by limiting the number of vehicles considered to build the RV-graph, limiting the size of groups in the RPV-graph, and limiting the number of decision variables ($|\epsilon_{PV}|$).

Algorithm 6: Batch-matching in horizon with pre-built routes algorithm

1 Initialize $V_{|R_v|\leq K}$ to V, R and R_O to empty lists, t to 0;
2 Set horizon to $[t, t+H]$, gather requests r emitted within the horizon and append new ones to R and R_O, gather AVs v with at most K requests in their service list to form $V_{|R_v|\leq K}$;
3 Build the RV-graph;
4 Build the RPV-graph by exploring the RV-graph. Each edge (P_i, v_j) of the RPV-graph carries the utility $u(P_i, v_j)$ which is the utility for v_j to be matched with the group of request P_i following plan $P^*_{R_v \cup P}$;
5 Solve the assignment problem between AVs and groups of requests in order to maximize the sum of utilities where one AV can be matched with at most one group;
6 Make permanent all assignments found in step 5 by setting v_j's plan to $P^*_{R_{v_j} \cup P_i}$ for all v_j so that x_{ij}=1;
7 Remove permanently matched requests from R_O;
8 Set $t = t + \alpha H$ and go to step 2;

B.4.1 Build the RV-graph

The RV-graph gives an overview of which requests might belong to the same plan and which AV might serve which request taking into account its current service plan. In the RV-graph:

- $r \in R_O$ and $r' \in R_O$ are connected if one virtual AV starting at the origin node of one of them could serve both while satisfying their time windows
- $r \in R_O$ and $v \in V_{|R_v|\leq K}$ are connected if v can serve r while satisfying the time windows of all requests in its current service plan (R_v) and the time window of r. The potential plan is $P^*_{R_v \cup \{r\}}$

B.4.2 Build the RPV-graph

The RPV-graph gives an overview of which AV might serve which group of requests. Let P be a group of requests. The RPV-graph contains:

- an edge between r and P if $r \in P$

- and edge between P and v if it exists a plan $P^*_{R_v \cup P}$ that satisfies the time windows of all requests in $R_v \cup P$. If several exists, the plan with the maximal utility is selected.

B.4.3 Assignment

The utility of a match between an AV v and a group of requests P is defined by:

$$u(P, v) = u(P^*_{R_v \cup P}) - u(P_{R_v}) \tag{B.1}$$

Let ϵ_{PV} be the set of $\{i, j\}$ for which edge (P_i, v_j) exists in the RPV-graph. Let x_{ij} be the binary decision variables of the assignment problem, where x_{ij} equals 1 if v_j is matched with P_i and 0 otherwise. The assignment problem becomes:

$$\begin{aligned}
\max_{x_{i,j}} \quad & \sum_{\{i,j\} \in \epsilon_{PV}} x_{ij} u(P_i, v_j) && \text{(B.2a)} \\
\text{subject to} \quad & x_{ij} \in \{0, 1\}, \forall \{i, j\} \in \epsilon_{PV} && \text{(B.2b)} \\
& \sum_{i \in I^V_j} x_{ij} \leq 1, \forall v_j \in V_{|R_v| \leq K} && \text{(B.2c)} \\
& \sum_{i \in I^R_k} \sum_{j \in I^P_i} x_{ij} \leq 1, \forall r_k \in R_O && \text{(B.2d)}
\end{aligned}$$

where I^V_j is the set of indices i for which edge (P_i, v_j) exists, I^R_k is the set of indices i for which edge (r_k, P_i) exists, I^P_i is the set of indices j for which edge (P_i, v_j) exists.

References

Agarwal, A. & Lämmel, G. (2016). Modeling seepage behavior of smaller vehicles in mixed traffic conditions using an agent based simulation. *Transportation in Developing Economies*, 2(2):8.

Akamatsu, T., Wada, K. & Hayashi, S. (2015). The corridor problem with discrete multiple bottlenecks. *Transportation Research Procedia*, 7:474–498.

Al-Kanj, L., Nascimento, J. & Powell, W. B. (2020). Approximate dynamic programming for planning a ride-hailing system using autonomous fleets of electric vehicles. *European Journal of Operational Research*, 284(3):1088–1106.

Aldaihani, M. & Dessouky, M. M. (2003). Hybrid scheduling methods for paratransit operations. *Computers & Industrial Engineering*, 45(1):75–96.

Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E. & Rus, D. (2017a). On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences*, 114(3):462–467.

Alonso-Mora, J., Wallar, A. & Rus, D. (2017b). Predictive routing for autonomous mobility-on-demand systems with ride-sharing. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3583–3590.

Ameli, M. (2019). *Heuristic Methods for Calculating Dynamic Traffic Assignment Simulation-based dynamic traffic assignment*. Ph.D. thesis, ENTPE - Université Gustave Eiffel.

André, J. & Vieira da Rocha, T. (2020). Émissions de gaz à effet de serre et de polluants atmosphériques en france. Tech. rep., Citepa.

Arnott, R., de Palma, A. & Lindsey, R. (1990). Economics of a bottleneck. *Journal of Urban Economics*, 27(1):111–130.

Axhausen, K. W., ETH Zürich, Horni, A., Nagel, K. & TU Berlin (2016). *The Multi-Agent Transport Simulation MATSim*. Ubiquity Press.

Axhausen, K. W., Meister, K., Balmer, M., Ciari, F., Horni, A., Rieser, M. & Waraich, R. A. (2010). Large-scale agent-based travel demand optimization applied to switzerland, including mode choice. *Working Paper / Institute for Transport Planning and Systems*. Doi: 10.3929/ETHZ-A-006100288, Publisher: ETH Zurich.

Ban, J., Wang, Y., Mackenzie, D., Guo, Q., Zhang, Y. & Angah, O. (2021). A multiscale simulation platform for connected and automated transportation systems. In *Transportation Research Board 100th Annual Meeting*.

Basciftci, B. & Van Hentenryck, P. (2023). Capturing travel mode adoption in designing on-demand multimodal transit systems. *Transportation Science*, 57(2):351–375.

Basu, R., Araldo, A., Akkinepally, A. P., Nahmias Biran, B. H., Basak, K., Seshadri, R., Deshmukh, N., Kumar, N., Azevedo, C. L. & Ben-Akiva, M. (2018). Automated mobility-on-demand vs. mass transit: A multi-modal activity-driven agent-based simulation approach. *Transportation Research Record: Journal of the Transportation Research Board*, 2672(8):608–618.

Becker, H., Becker, F., Abe, R., Bekhor, S., Belgiawan, P. F., Compostella, J., Frazzoli, E., Fulton, L. M., Guggisberg Bicudo, D., Murthy Gurumurthy, K., Hensher, D. A., Joubert, J. W., Kockelman, K. M., Kröger, L., Le Vine, S., Malik, J., Marczuk, K., Ashari Nasution, R., Rich, J., Papu Carrone, A., Shen, D., Shiftan, Y., Tirachini, A., Wong, Y. Z., Zhang, M., Bösch, P. M. & Axhausen, K. W. (2020). Impact of vehicle automation and electric propulsion on production costs for mobility services worldwide. *Transportation Research Part A: Policy and Practice*, 138:105–126.

Ben-Dor, G., Ben-Elia, E. & Benenson, I. (2021). Population downscaling in multi-agent transportation simulations: A review and case study. *Simulation Modelling Practice and Theory*, 108:102233.

Ben-Dor, G., Ogulenko, A., Klein, I. & Benenson, I. (2022). Modal shift and shared automated demand-responsive transport: A case study of jerusalem. *Procedia Computer Science*, 201:581–586.

Berbeglia, G., Cordeau, J.-F. & Laporte, G. (2010). Dynamic pickup and delivery problems. *European Journal of Operational Research*, 202(1):8–15.

Bimpikis, K., Candogan, O. & Saban, D. (2019). Spatial pricing in ride-sharing networks. *Operations Research*, 67(3):744–769.

Bonnel, P. (2004). *Prévoir la demande de transport*. Presses de l'école nationale des ponts et chaussées.

Bourgeois, G. (2017). Modélisation des véhicules autonomes partagés: état de l'art et interrogations. Tech. rep.

Braekers, K., Caris, A. & Janssens, G. K. (2014). Exact and meta-heuristic approach for a general heterogeneous dial-a-ride problem with multiple depots. *Transportation Research Part B: Methodological*, 67:166–186.

Buisson, C. & Ladier, C. (2009). Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *Transportation Research Record: Journal of the Transportation Research Board*, 2124(1):127–136.

Carron, A., Seccamonte, F., Ruch, C., Frazzoli, E. & Zeilinger, M. N. (2021). Scalable model predictive control for autonomous mobility-on-demand systems. *IEEE Transactions on Control Systems Technology*, 29(2):635–644.

CEBR (2014). The future economic and environmental costs of gridlock in 2030: An assessment of the direct and indirect economic and environmental costs of idling in road traffic congestion to households in the UK, france, germany and the USA. Tech. rep., Center of Economics and Business Research.

Cerema (2021). Prise de la compétence d'organisation des mobilités par les communautés de communes : panorama au 15 juillet 2021. Url: https://www.cerema.fr/fr/actualites/prise-competence-organisation-mobilites-communautes-communes.

Chiabaut, N., Comemale, A. & Laval, J. A. (2018). Analysis of dynamic multimodal user equilibrium for the single destination problem. *Transportation Research Record: Journal of the Transportation Research Board*. Unpublished.

Chugh, T., Jin, Y., Miettinen, K., Hakanen, J. & Sindhya, K. (2018). A surrogate-assisted reference vector guided evolutionary algorithm for computationally expensive many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 22(1):129–142.

Cordeau, J.-F. & Laporte, G. (2007). The dial-a-ride problem: models and algorithms. *Annals of Operations Research*, 153(1):29–46.

Cubillos, C., Urra, E. & Rodríguez, N. (2009). Application of genetic algorithms for the DARPTW problem. *International Journal of Computers Communications & Control*, 4(2):127–136.

Currie, G. (2005). The demand performance of bus rapid transit. *Journal of Public Transportation*, 8(1):41–55.

Daganzo, C. F. (2010). Structure of competitive transit networks. *Transportation Research Part B: Methodological*, 44(4):434–446.

Dandl, F., Engelhardt, R., Hyland, M., Tilg, G., Bogenberger, K. & Mahmassani, H. S. (2021). Regulating mobility-on-demand services: Tri-level model and bayesian optimization solution approach. *Transportation Research Part C: Emerging Technologies*, 125:103075.

Daulton, S., Balandat, M. & Bakshy, E. (2021). Parallel bayesian optimization of multiple noisy objectives with expected hypervolume improvement. Doi: 10.48550/ARXIV.2105.08195.

Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197.

Desrosiers, J., Dumas, Y. & Soumis, F. (1986). A dynamic programming solution of the large-scale single-vehicle dial-a-ride problem with time windows. *American Journal of Mathematical and Management Sciences*, 6(3):301–325.

Dia, H. & Javanshour, F. (2017). Autonomous shared mobility-on-demand: Melbourne pilot simulation study. *Transportation Research Procedia*, 22:285–296.

Diallo, A. O., Moliner, M., Manout, O. & Ciari, F. (2023). Implications of three autonomous mobility scenarios: A comparison between lyon, france and montreal, canada. Doi: 10.2139/ssrn.4383042.

Diana, M. & Dessouky, M. M. (2004). A new regret insertion heuristic for solving large-scale dial-a-ride problems with time windows. *Transportation Research Part B: Methodological*, 38(6):539–557.

Dong, X. (2020). Trade uber for the bus?: An investigation of individual willingness to use ride-hail versus transit. *Journal of the American Planning Association*, 86(2):222–235.

Dowling, R. & McGuirk, P. (2022). Autonomous vehicle experiments and the city. *Urban Geography*, 43(3):409–426.

Emmerich, M., Giannakoglou, K. & Naujoks, B. (2006). Single- and multiobjective evolutionary optimization assisted by gaussian random field metamodels. *IEEE Transactions on Evolutionary Computation*, 10(4):421–439.

Erhardt, G. D., Hoque, J. M., Goyal, V., Berrebi, S., Brakewood, C. & Watkins, K. E. (2022a). Why has public transit ridership declined in the united states? *Transportation Research Part A: Policy and Practice*, 161:68–87.

Erhardt, G. D., Mucci, R. A., Cooper, D., Sana, B., Chen, M. & Castiglione, J. (2022b). Do transportation network companies increase or decrease transit ridership? empirical evidence from san francisco. *Transportation*, 49(2):313–342.

Erhardt, G. D., Roy, S., Cooper, D., Sana, B., Chen, M. & Castiglione, J. (2019). Do transportation network companies decrease or increase congestion? *Science Advances*, 5(5):eaau2670.

Estrada, M., Roca-Riu, M., Badia, H., Robusté, F. & Daganzo, C. (2011). Design and implementation of efficient transit networks: Procedure, case study and validity test. *Procedia - Social and Behavioral Sciences*, 17:113–135.

Fagnant, D. J. & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, 77:167–181.

Fagnant, D. J. & Kockelman, K. M. (2014). The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. *Transportation Research Part C: Emerging Technologies*, 40:1–13.

Fagnant, D. J. & Kockelman, K. M. (2018). Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in austin, texas. *Transportation*, 45(1):143–158.

Fagnant, D. J., Kockelman, K. M. & Bansal, P. (2016). Operations of shared autonomous vehicle fleet for austin, texas, market. *Transportation Research Record: Journal of the Transportation Research Board*, 2563(1):98–106.

Fakhrmoosavi, F., Kamjoo, E., Kavianipour, M., Zockaie, A., Talebpour, A. & Mittal, A. (2022). A stochastic framework using bayesian optimization algorithm to assess the network-level societal impacts of connected and autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, 139:103663.

Flötteröd, G. & Rohde, J. (2011). Operational macroscopic modeling of complex urban road intersections. *Transportation Research Part B: Methodological*, 45(6):903–922.

Fosgerau, M. & de Palma, A. (2012). Congestion in a city with a central bottleneck. *Journal of Urban Economics*, 71(3):269–277.

Fraedrich, E., Heinrichs, D., Bahamonde-Birke, F. J. & Cyganski, R. (2019). Autonomous driving, the built environment and policy implications. *Transportation Research Part A: Policy and Practice*, 122:162–172.

Frei, C., Hyland, M. & Mahmassani, H. S. (2017). Flexing service schedules: Assessing the potential for demand-adaptive hybrid transit via a stated preference approach. *Transportation Research Part C: Emerging Technologies*, 76:71–89.

Fu, Q., Liu, R. & Hess, S. (2012). A review on transit assignment modelling approaches to congested networks: A new perspective. *Procedia - Social and Behavioral Sciences*, 54:1145–1155.

Gallotti, R. & Barthelemy, M. (2015). Anatomy and efficiency of urban multimodal mobility. *Scientific Reports*, 4(1):6911.

Galuzio, P. P., De Vasconcelos Segundo, E. H., Coelho, L. D. S. & Mariani, V. C. (2020). MOBOpt — multi-objective bayesian optimization. *SoftwareX*, 12:100520.

Gao, J. & Li, S. (2023). Regulating for-hire autonomous vehicles for an equitable multimodal transportation network. Doi: arXiv:2301.05798.

Garcia-Martinez, A., Cascajo, R., Jara-Diaz, S. R., Chowdhury, S. & Monzon, A. (2018). Transfer penalties in multimodal public transport networks. *Transportation Research Part A: Policy and Practice*, 114:52–66.

Gawron, J. H., Keoleian, G. A., De Kleine, R. D., Wallington, T. J. & Kim, H. C. (2018). Life cycle assessment of connected and automated vehicles: Sensing and computing subsystem and vehicle level effects. *Environmental Science & Technology*, 52(5):3249–3256.

George, D. (2012). *Stochastic modeling and decentralized control policies for large-scale vehicle sharing systems via closed queueing networks*. Ph.D. thesis, Ohio State University.

Geroliminis, N. & Daganzo, C. F. (2007). Macroscopic modeling of traffic in cities. In *Transportation Research Board 86th Annual Meeting*.

Gipps, P. (1981). A behavioural car-following model for computer simulation. *Transportation Research Part B: Methodological*, 15(2):105–111.

Gipps, P. (1986). A model for the structure of lane-changing decisions. *Transportation Research Part B: Methodological*, 20(5):403–414.

Gueriau, M., Cugurullo, F., Acheampong, R. A. & Dusparic, I. (2020). Shared autonomous mobility on demand: A learning-based approach and its performance in the presence of traffic congestion. *IEEE Intelligent Transportation Systems Magazine*, 12(4):208–218.

Gueriau, M. & Dusparic, I. (2020). Quantifying the impact of connected and autonomous vehicles on traffic efficiency and safety in mixed traffic. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–8.

Gurumurthy, K. M., Auld, J. & Kockelman, K. (2021). A system of shared autonomous vehicles for chicago: Understanding the effects of geofencing the service. *Journal of Transport*

and Land Use, 14(1):933–948.

Gurumurthy, K. M. & Kockelman, K. M. (2022). Dynamic ride-sharing impacts of greater trip demand and aggregation at stops in shared autonomous vehicle systems. *Transportation Research Part A: Policy and Practice*, 160:114–125.

Gurumurthy, K. M., Kockelman, K. M. & Zuniga-Garcia, N. (2020). First-mile-last-mile collector-distributor system using shared autonomous mobility. *Transportation Research Record: Journal of the Transportation Research Board*.

Hall, J. D., Palsson, C. & Price, J. (2018). Is uber a substitute or complement for public transit? *Journal of Urban Economics*, 108:36–50.

Hebbal, A., Brevault, L., Balesdent, M., Talbi, E.-G. & Melab, N. (2019). Multi-objective optimization using deep gaussian processes: Application to aerospace vehicle design. In *AIAA Scitech 2019 Forum*. American Institute of Aeronautics and Astronautics.

Ho, S. C., Szeto, W., Kuo, Y.-H., Leung, J. M., Petering, M. & Tou, T. W. (2018). A survey of dial-a-ride problems: Literature review and recent developments. *Transportation Research Part B: Methodological*, 111:395–421.

Hou, Y., Young, S. E., Garikapati, V., Chen, Y. & Zhu, L. (2018). Initial assessment and modeling framework development for automated mobility districts. In *ITS World Congress 2017: Integrated Mobility Driving Smart Cities*.

Hryhoryeva, M. & Leclercq, L. (2023). Competition in ride-hailing service operations: Impacts on travel distances and service performance. *Transportation Research Record: Journal of the Transportation Research Board*.

Huang, Y., Kockelman, K. M., Garikapati, V., Zhu, L. & Young, S. (2021). Use of shared automated vehicles for first-mile last-mile service: Micro-simulation of rail-transit connections in austin, texas. *Transportation Research Record: Journal of the Transportation Research Board*, 2675(2):135–149.

Huo, J., Liu, Z., Chen, J., Cheng, Q. & Meng, Q. (2023). Bayesian optimization for congestion pricing problems: A general framework and its instability. *Transportation Research Part B: Methodological*, 169:1–28.

Hyland, M. & Mahmassani, H. S. (2018). Dynamic autonomous vehicle fleet operations: Optimization-based strategies to assign AVs to immediate traveler demand requests. *Transportation Research Part C: Emerging Technologies*, 92:278–297.

Hörl, S., Becker, F. & Axhausen, K. W. (2021). Simulation of price, customer behaviour and system impact for a cost-covering automated taxi system in zurich. *Transportation Research Part C: Emerging Technologies*, 123:102974.

Hörl, S., Ruch, C., Becker, F., Frazzoli, E. & Axhausen, K. (2019). Fleet operational policies for automated mobility: A simulation assessment for zurich. *Transportation Research Part C: Emerging Technologies*, 102:20–31.

Iryo, T. (2013). Properties of dynamic user equilibrium solution: existence, uniqueness, stability, and robust solution methodology. *Transportmetrica B: Transport Dynamics*, 1(1):52–67.

Javanshour, F., Dia, H., Duncan, G., Abduljabbar, R. & Liyanage, S. (2022). Performance evaluation of station-based autonomous on-demand car-sharing systems. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):7721–7732.

Jin, J., Zhou, M., Zhang, W., Li, M., Guo, Z., Qin, Z., Jiao, Y., Tang, X., Wang, C., Wang, J., Wu, G. & Ye, J. (2019). CoRide: Joint order dispatching and fleet management for multi-scale ride-hailing platforms. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 1983–1992.

Jin, K., Wang, W., Hua, X. & Zhou, W. (2020). Reinforcement learning for optimizing driving policies on cruising taxis services. *Sustainability*, 12(21):8883.

Jin, W.-L. (2015). Point queue models: A unified approach. *Transportation Research Part B: Methodological*, 77:1–16.

Jung, J., Jayakrishnan, R. & Park, J. Y. (2016). Dynamic shared-taxi dispatch algorithm with hybrid-simulated annealing. *Computer-Aided Civil and Infrastructure Engineering*, 31(4):275–291.

Ke, J., Xiao, F., Yang, H. & Ye, J. (2019). Optimizing online matching for ride-sourcing services with multi-agent deep reinforcement learning. Doi: 10.48550/ARXIV.1902.06228.

Ke, J., Zhu, Z., Yang, H. & He, Q. (2021). Equilibrium analyses and operational designs of a coupled market with substitutive and complementary ride-sourcing services to public transits. *Transportation Research Part E: Logistics and Transportation Review*, 148:102236.

Keane, A. J. (2006). Statistical improvement criteria for use in multiobjective design optimization. *AIAA Journal*, 44(4):879–891.

Kenworthy, J. (2003). Transport energy use and greenhouse gases in urban passengertransport systems: A study of 84 global cities.

Kerry, C. & Karsten, J. (2017). Gauging investment in self-driving cars. Url: https://www.brookings.edu/research/gauging-investment-in-self-driving-cars/.

Khemiri, R., Naija, M. & Exposito, E. (2022). Dispatching and rebalancing for ride-sharing autonomous mobility-on-demand systems based on a fuzzy multi-criteria approach. *Soft Computing*.

Knowles, J. (2006). ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Transactions on Evolutionary Computation*, 10(1):50–66.

Knowles, J. & Corne, D. (1999). The pareto archived evolution strategy: a new baseline algorithm for pareto multiobjective optimisation. In *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)*, pages 98–105.

Kolarova, V., Steck, F., Cyganski, R. & Trommer, S. (2018). Estimation of the value of time for automated driving using revealed and stated preference methods. *Transportation Research Procedia*, 31:35–46.

Kumar, P. & Khani, A. (2022). Planning of integrated mobility-on-demand and urban transit networks. *Transportation Research Part A: Policy and Practice*, 166:499–521.

Lang, N. S., Ruessmann, M., Collie, B. E., Wegscheider, A. K., Moavenzadeh, J. & Loane, M. (2018). Reshaping urban mobility with autonomous vehicles: Lessons from the city of boston. Tech. rep., World Economic Forum.

Laval, J. A. (2009). Graphical solution and continuum approximation for the single destination dynamic user equilibrium problem. *Transportation Research Part B: Methodological*, 43(1):108–118.

Lejri, D., Jeanneret, B., Bécarie, C. & Trigui, R. (2023). Vehicle fleet electrification: electric energy consumption assessment at the scale of an urban area. In *25th International Transport and Air Pollution Conference (TAP)*.

Lejri, D. & Leclercq, L. (2020). Are average speed emission functions scale-free? *Atmospheric Environment*, 224:117324.

Levin, M. W., Kockelman, K. M., Boyles, S. D. & Li, T. (2017). A general framework for modeling shared autonomous vehicles with dynamic network-loading and dynamic ride-sharing application. *Computers, Environment and Urban Systems*, 64:373–383.

Li, M., Chen, T. & Yao, X. (2022). How to evaluate solutions in pareto-based search-based software engineering: A critical review and methodological guidance. *IEEE Transactions on Software Engineering*, 48(5):1771–1799.

Li, S., Tavafoghi, H., Poolla, K. & Varaiya, P. (2019). Regulating TNCs: Should uber and lyft set their own rules? *Transportation Research Part B: Methodological*, 129:193–225.

Li, S., Yang, H., Poolla, K. & Varaiya, P. (2021). Spatial pricing in ride-sourcing markets under a congestion charge. *Transportation Research Part B: Methodological*, 152:18–45.

Li, X. (2018). *Fairness in Transportation System*. Ph.D. thesis, UNSW Sydney.

Li, X. & Quadrifoglio, L. (2010). Feeder transit services: Choosing between fixed and demand responsive policy. *Transportation Research Part C: Emerging Technologies*, 18(5):770–780.

Lighthill, M. J. & Whitham, G. B. (1955). On kinematic waves. II. a theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 229(1178):317–345.

Lin, K., Zhao, R., Xu, Z. & Zhou, J. (2018). Efficient large-scale fleet management via multi-agent deep reinforcement learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1774–1783.

Liu, R., Jiang, Y. & Azevedo, C. L. (2021). Bayesian optimization of area-based road pricing. In *2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pages 1–6.

Liu, W. (2018). An equilibrium analysis of commuter parking in the era of autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, 92:191–207.

Liu, Y., Bansal, P., Daziano, R. & Samaranayake, S. (2019). A framework to integrate mode choice in the design of mobility-on-demand systems. *Transportation Research Part C: Emerging Technologies*, 105:648–665.

Liu, Y. & Ouyang, Y. (2021). Mobility service design via joint optimization of transit networks and demand-responsive services. *Transportation Research Part B: Methodological*, 151:22–41.

Liu, Z., Li, J. & Wu, K. (2020). Context-aware taxi dispatching at city-scale using deep reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–14.

Lokhandwala, M. & Cai, H. (2018). Dynamic ride sharing using traditional taxis and shared autonomous taxis: A case study of NYC. *Transportation Research Part C: Emerging Technologies*, 97:45–60.

Lopez, P. A., Wiessner, E., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flotterod, Y.-P., Hilbrich, R., Lucken, L., Rummel, J. & Wagner, P. (2018). Microscopic traffic simulation using SUMO. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 2575–2582.

Low, K. Y. A., Vissol-Gaudin, E., Lim, Y. F. & Hippalgaonkar, K. (2022). Bayesian vs evolutionary optimisation in exploring pareto fronts for materials discovery. Doi: 10.36227/techrxiv.21154537.v1.

Lu, Q., Tettamanti, T., Hörcher, D. & Varga, I. (2020). The impact of autonomous vehicles on urban traffic network capacity: an experimental analysis by microscopic traffic simulation. *Transportation Letters*, 12(8):540–549.

Luo, Y. & Schonfeld, P. (2011). Online rejected-reinsertion heuristics for dynamic multivehicle dial-a-ride problem. *Transportation Research Record: Journal of the Transportation Research Board*, 2218(1):59–67.

Ma, J., Li, X., Zhou, F. & Hao, W. (2017). Designing optimal autonomous vehicle sharing and reservation systems: A linear programming approach. *Transportation Research Part C: Emerging Technologies*, 84:124–141.

Maciejewski, M. (2015). Benchmarking minimum passenger waiting time in online taxi dispatching with exact offline optimization methods. *Archive of Transport*, 30(2):67–75.

Maciejewski, M., Bischoff, J. & Nagel, K. (2016). An assignment-based approach to efficient real-time city-scale taxi dispatching. *IEEE Intelligent Systems*, 31(1):68–77.

Maerivoet, S. & De Moor, B. (2005). Transportation planning and traffic flow models. Doi: arXiv:physics/0507127.

Mahmassani, H. S., Williams, J. C. & Herman, R. (1984). Investigation of network-level traffic flow relationships: some simulation results. *Transportation Research Record: Journal of the Transportation Research Board*, 971:121–130.

Mao, C., Liu, Y. & Shen, Z.-J. M. (2020). Dispatch of autonomous vehicles for taxi services: A deep reinforcement learning approach. *Transportation Research Part C: Emerging Technologies*, 115:102626.

Mariotte, G. (2018). *Dynamic Modeling of Large-Scale Urban Transportation Systems*. Ph.D. thesis, ENTPE - Université Gustave Eiffel.

Masoud, N. & Jayakrishnan, R. (2017). Autonomous or driver-less vehicles: Implemen-

tation strategies and operational concerns. *Transportation Research Part E: Logistics and Transportation Review*, 108:179–194.

Melchiorri, M., Florczyk, A., Freire, S., Schiavina, M., Pesaresi, M. & Kemper, T. (2018). Unveiling 25 years of planetary urbanization with remote sensing: Perspectives from the global human settlement layer. *Remote Sensing*, 10(5):768.

Meyer, L. H. & Roser, D. (2009). Enough for the future. In *Intergenerational Justice*, pages 219–248. Oxford University PressOxford.

Milakis, D., van Arem, B. & van Wee, B. (2017). Policy and society related implications of automated driving: A review of literature and directions for future research. *Journal of Intelligent Transportation Systems*, 21(4):324–348.

Militão, A. M. & Tirachini, A. (2021). Optimal fleet size for a shared demand-responsive transport system with human-driven vs automated vehicles: A total cost minimization approach. *Transportation Research Part A: Policy and Practice*, 151:52–80.

Mo, B., Cao, Z., Zhang, H., Shen, Y. & Zhao, J. (2021). Competition between shared autonomous vehicles and public transit: A case study in singapore. *Transportation Research Part C: Emerging Technologies*, 127:103058.

Moreno, A. T., Michalski, A., Llorca, C. & Moeckel, R. (2018). Shared autonomous vehicles effect on vehicle-km traveled and average trip duration. *Journal of Advanced Transportation*, 2018:1–10.

Mounce, R. (2009). Existence of equilibrium in a continuous dynamic queueing model for traffic networks with responsive signal control. In *Transportation and Traffic Theory 2009: Golden Jubilee*, pages 327–344. Springer, Boston, MA.

Narayan, J., Cats, O., Van Oort, N. & Hoogendoorn, S. (2019). Does ride-sourcing absorb the demand for car and public transport in amsterdam? In *2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pages 1–7.

Narayanan, S., Chaniotakis, E. & Antoniou, C. (2020). Shared autonomous vehicle services: A comprehensive review. *Transportation Research Part C: Emerging Technologies*, 111:255–293.

Nelson, E. & Sadowsky, N. (2019). Estimating the impact of ride-hailing app company entry on public transportation use in major US urban areas. *The B.E. Journal of Economic Analysis & Policy*, 19(1).

Newell, G. (1993). A simplified theory of kinematic waves in highway traffic, part i: General theory. *Transportation Research Part B: Methodological*, 27(4):281–287.

Nicolas, J., Verry, D. & Longuar, Z. (2012). Évolutions récentes des émissions de CO2 liées à la mobilité des français : analyser les dynamiques à l'œuvre grâce aux enquêtes nationales transports de 1994 et 2008. *Economie et statistique*, 457(1):161–183.

Nourinejad, M. & Ramezani, M. (2020). Ride-sourcing modeling and pricing in non-equilibrium two-sided markets. *Transportation Research Part B: Methodological*, 132:340–357.

Paipuri, M., Barmpounakis, E., Geroliminis, N. & Leclercq, L. (2021). Empirical observations of multi-modal network-level models: Insights from the pNEUMA experiment. *Transportation Research Part C: Emerging Technologies*, 131:103300.

Parragh, S. N., Doerner, K. F., Hartl, R. F. & Gandibleux, X. (2009). A heuristic two-phase solution approach for the multi-objective dial-a-ride problem. *Networks*, 54(4):227–242.

Pavone, M. (2010). *Dynamic vehicle routing for robotic networks*. Ph.D. thesis, Massachusetts Institute of Technology.

Pavone, M. (2015). Autonomous mobility-on-demand systems for future urban mobility. In *Autonomes Fahren*, pages 399–416. Springer Berlin Heidelberg.

Pavone, M., Bisnik, N., Frazzoli, E. & Isler, V. (2009). A stochastic and dynamic vehicle routing problem with time windows and customer impatience. *Mobile Networks and Applications*, 14(3):350–364.

Pavone, M., Smith, S. L., Frazzoli, E. & Rus, D. (2012). Robotic load balancing for mobility-on-demand systems. *The International Journal of Robotics Research*, 31(7):839–854.

Picheny, V. (2015). Multiobjective optimization using gaussian process emulators via stepwise uncertainty reduction. *Statistics and Computing*, 25(6):1265–1280.

Picheny, V., Binois, M. & Habbal, A. (2019). A bayesian optimization approach to find nash equilibria. *Journal of Global Optimization*, 73:171–192.

Pimenta, V., Quilliot, A., Toussaint, H. & Vigo, D. (2017). Models and algorithms for reliability-oriented dial-a-ride with autonomous electric vehicles. *European Journal of Operational Research*, 257(2):601–613.

Pinto, H. K., Hyland, M. F., Mahmassani, H. S. & Verbas, I. O. (2020). Joint design of multimodal transit networks and shared autonomous mobility fleets. *Transportation Research Part C: Emerging Technologies*, 113:2–20.

Ponweiser, W., Wagner, T., Biermann, D. & Vincze, M. (2008). Multiobjective optimization on a limited budget of evaluations using model-assisted s-metric selection. In *Parallel Problem Solving from Nature – PPSN X*, vol. 5199, pages 784–794. Springer Berlin Heidelberg.

Priem, R. (2020). *Optimisation bayésienne sous contraintes et en grande dimension appliquée à la conception avion avant projet*. Ph.D. thesis, Toulouse, ISAE.

Psaraftis, H. (1988). Dynamic vehicle routing problems. *Vehicle routing: Methods and studies*, 16:223–248.

Pöhler, L. D., Asami, Y. & Oguchi, T. (2019). Urban land use policies for efficient autonomous on-demand transportation – a case study on the japanese island of izu oshima. *International Journal of Transport Development and Integration*, 3(2):152–165.

Rayle, L., Dai, D., Chan, N., Cervero, R. & Shaheen, S. (2016). Just a better taxi? a survey-based comparison of taxis, transit, and ridesourcing services in san francisco. *Transport Policy*, 45:168–178.

Reck, D. J. & Axhausen, K. W. (2019). Subsidized ridesourcing for the first/last mile: How

valuable for whom? *European Journal of Transport and Infrastructure Research*, 20(4):59–77.

Reid, D. (1996). Genetic algorithms in constrained optimization. *Mathematical and Computer Modelling*, 23(5):87–111.

Richards, P. I. (1956). Shock waves on the highway. *Operations Research*, 4(1):42–51.

Riedmaier, S., Schneider, D., Watzenig, D., Diermeyer, F. & Schick, B. (2020). Model validation and scenario selection for virtual-based homologation of automated vehicles. *Applied Sciences*, 11(1):35.

Rieser, M., Grether, D. & Nagel, K. (2009). Adding mode choice to multiagent transport simulation. *Transportation Research Record: Journal of the Transportation Research Board*, 2132(1):50–58.

Rifki, O., Chiabaut, N. & Nicolas, J. (2021). Design of an autonomous ride-sharing service through a graph embedding integrated 2 dial-a-ride problem: Application to the last-mile transit in lyon city. In *Transportation Research Board 100th Annual Meeting*.

Ropke, S. & Pisinger, D. (2006). An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. *Transportation Science*, 40(4):455–472.

Räth, Y. M., Balać, M., Hörl, S. & Axhausen, K. W. (2021). System level policies for automated transit on-demand services: A case study for zurich, switzerland. Doi: 10.3929/ethz-b-000474208, Publisher: ETH Zurich.

Salazar, M., Lanzetti, N., Rossi, F., Schiffer, M. & Pavone, M. (2020). Intermodal autonomous mobility-on-demand. *IEEE Transactions on Intelligent Transportation Systems*, 21(9):3946–3960.

Salazar, M., Rossi, F., Schiffer, M., Onder, C. H. & Pavone, M. (2018). On the interaction between autonomous mobility-on-demand and public transportation systems. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 2262–2269.

Salazar, M., Tsao, M., Aguiar, I., Schiffer, M. & Pavone, M. (2019). A congestion-aware routing scheme for autonomous mobility-on-demand systems. In *2019 18th European Control Conference (ECC)*, pages 3040–3046.

Saujot, M., Brimont, L. & Sartor, O. (2018). Mettons la mobilité autonome sur la voie du développement durable. *Studies Iddri*, 18(2).

Savelsbergh, M. W. P. & Sol, M. (1995). The general pickup and delivery problem. *Transportation Science*, 29(1):17–29.

Scheltes, A. & de Almeida Correia, G. H. (2017). Exploring the use of automated vehicles as last mile connection of train trips through an agent-based simulation model: An application to delft, netherlands. *International Journal of Transportation Science and Technology*, 6(1):28–41.

Schrank, D., Eisele, B. & Lomax, T. (2019). Urban mobility report 2019. Tech. rep., Texas Transportation Institute.

Sean Qian, Z. & Michael Zhang, H. (2011). Modeling multi-modal morning commute in a one-to-one corridor network. *Transportation Research Part C: Emerging Technologies*,

19(2):254–269.

Sha, D., Ozbay, K. & Ding, Y. (2020). Applying bayesian optimization for calibration of transportation simulation models. *Transportation Research Record: Journal of the Transportation Research Board*, 2674(10):215–228.

Shaheen, S. & Chan, N. (2016). Mobility and the sharing economy: Potential to facilitate the first- and last-mile public transit connections. *Built Environment*, 42(4):573–588.

Shan, A., Hoang, N. H., An, K. & Vu, H. L. (2021). A framework for railway transit network design with first-mile shared autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, 130:103223.

Sheffi, Y. & Powell, W. B. (1982). An algorithm for the equilibrium assignment problem with random link times. *Networks*, 12(2):191–207.

Shen, C.-W. & Quadrifoglio, L. (2013). Evaluating centralized versus decentralized zoning strategies for metropolitan ADA paratransit services. *Journal of Transportation Engineering*, 139(5):524–532.

Shen, Y., Zhang, H. & Zhao, J. (2018). Integrating shared autonomous vehicle in public transportation system: A supply-side simulation of the first-mile service in singapore. *Transportation Research Part A: Policy and Practice*, 113:125–136.

Shinkyu Jeong & Obayashi, S. (2005). Efficient global optimization (EGO) for multi-objective problem and data mining. In *2005 IEEE Congress on Evolutionary Computation*, vol. 3, pages 2138–2145.

Shou, Z. & Di, X. (2020). Reward design for driver repositioning using multi-agent reinforcement learning. *Transportation Research Part C: Emerging Technologies*, 119:102738.

Simoni, M. D., Kockelman, K. M., Gurumurthy, K. M. & Bischoff, J. (2019). Congestion pricing in a world of self-driving vehicles: An analysis of different strategies in alternative future scenarios. *Transportation Research Part C: Emerging Technologies*, 98:167–185.

Smith, M. & Ghali, M. (1990). Dynamic traffic assignment and dynamic traffic control. In *Transportation and Traffic Theory: Proceedings of the Eleventh International Symposium on Transportation and Traffic Theory*, pages 273–290. M. Koshi.

Stein, G. (2021). The impact of autonomous vehicles on urban land use patterns. *Florida State University Law Review*, 48:193.

Sun, H., Wang, H. & Wan, Z. (2019). Model and analysis of labor supply for ride-sharing platforms in the presence of sample self-selection and endogeneity. *Transportation Research Part B: Methodological*, 125:76–93.

Tampere, C., Viti, F. & Immers, L. (2010). *New Developments in Transport Planning*. Edward Elgar Publishing.

Tay, T. & Osorio, C. (2022). Bayesian optimization techniques for high-dimensional simulation-based transportation problems. *Transportation Research Part B: Methodological*, 164:210–243.

Titos, G., Lyamani, H., Drinovec, L., Olmo, F., Močnik, G. & Alados-Arboledas, L. (2015).

Evaluation of the impact of transportation changes on air quality. *Atmospheric Environment*, 114:19–31.

United Nations (2019). World urbanization prospects: The 2018 revision. Tech. rep., United Nations, Department of Economic and Social Affairs, Population Division.

Université Gustave Eiffel (2019). Le projet ENA: Expérimentation de navettes autonomes. Dossier de presse, Université Gustave Eiffel.

Urba Lyon (2018). Pratiques de déplacements sur les bassins de vie de scot de l'agglomération lyonnaise: enquête déplacements 2015. Tech. rep., SYTRAL. Url: https://www.scot-agglolyon.fr/les-documents-du-scot/.

Vallée, S., Oulamara, A. & Cherif-Khettaf, W. R. (2017). Maximizing the number of served requests in an online shared transport system by solving a dynamic DARP. In *Computational Logistics*, vol. 10572, pages 64–78.

Veve, C. (2021). *Détection de zones spatio-temporelles à fort potentiel pour contribuer à la mise en place de services de mobilité partagée et de microtransit*. Ph.D. thesis, ENTPE - Université Gustave Eiffel.

Vickrey, W. S. (1969). Congestion theory and transport investment. *The American Economic Review*, 59(2):251–260.

Vignon, D. A., Yin, Y. & Ke, J. (2021). Regulating ridesourcing services with product differentiation and congestion externality. *Transportation Research Part C: Emerging Technologies*, 127:103088.

Vinot, E., Scordia, J., Trigui, R., Jeanneret, B. & Badin, F. (2008). Model simulation, validation and case study of the 2004 THS of toyota prius. *International Journal of Vehicle Systems Modelling and Testing*, 3(3):139.

Vuchic, V. R. (1969). Rapid transit interstation spacings for maximum number of passengers. *Transportation Science*, 3(3):214–232.

Vuchic, V. R. & Newell, G. F. (1968). Rapid transit interstation spacings for minimum travel time. *Transportation Science*, 2(4):303–339.

Wagstaff, A., Paci, P. & van Doorslaer, E. (1991). On the measurement of inequalities in health. *Social Science & Medicine*, 33(5):545–557.

Wang, H., Xu, H. & Zhang, Z. (2023). High-dimensional multi-objective bayesian optimization with block coordinate updates: Case studies in intelligent transportation system. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–12.

Wardrop, J. G. (1952). Road paper. some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1(3):325–362.

Wei, K., Vaze, V. & Jacquillat, A. (2022). Transit planning optimization under ride-hailing competition and traffic congestion. *Transportation Science*, 56(3):725–749.

Wen, J., Chen, Y. X., Nassir, N. & Zhao, J. (2018). Transit-oriented autonomous vehicle operation with integrated demand-supply interaction. *Transportation Research Part C: Emerging Technologies*, 97:216–234.

Wirasinghe, S. C., Hurdle, V. F. & Newell, G. F. (1977). Optimal parameters for a coordinated rail and bus transit system. *Transportation Science*, 11(4):359–374.

Wollenstein-Betech, S., Salazar, M., Houshmand, A., Pavone, M., Paschalidis, I. C. & Cassandras, C. G. (2022). Routing and rebalancing intermodal autonomous mobility-on-demand systems in mixed traffic. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):12263–12275.

Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H. & Deng, S.-H. (2019). Hyperparameter optimization for machine learning models based on bayesian optimizationb. *Journal of Electronic Science and Technology*, 17(1):26–40.

Wu, W. & Huang, H. (2014). Equilibrium and modal split in a competitive highway/transit system under different road-use pricing strategies. *Journal of Transport Economics and Policy (JTEP)*, 48.

Yao, F., Zhu, J., Yu, J., Chen, C. & Chen, X. M. (2020). Hybrid operations of human driving vehicles and automated vehicles with data-driven agent-based simulation. *Transportation Research Part D: Transport and Environment*, 86:102469.

Yap, M. D., Correia, G. & van Arem, B. (2016). Preferences of travellers for using automated vehicles as last mile public transport of multimodal train trips. *Transportation Research Part A: Policy and Practice*, 94:1–16.

Yi, Z. & Smart, J. (2021). A framework for integrated dispatching and charging management of an autonomous electric vehicle ride-hailing fleet. *Transportation Research Part D: Transport and Environment*, 95:102822.

Yu, X., Gao, S., Hu, X. & Park, H. (2019). A markov decision process approach to vacant taxi routing with e-hailing. *Transportation Research Part B: Methodological*, 121:114–134.

Zardini, G., Lanzetti, N., Pavone, M. & Frazzoli, E. (2022). Analysis and control of autonomous mobility-on-demand systems. *Annual Review of Control, Robotics, and Autonomous Systems*, 5(1):633–658.

Zgraggen, J., Tsao, M., Salazar, M., Schiffer, M. & Pavone, M. (2019). A model predictive control scheme for intermodal autonomous mobility-on-demand. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 1953–1960.

Zha, L., Yin, Y. & Du, Y. (2018). Surge pricing and labor supply in the ride-sourcing market. *Transportation Research Part B: Methodological*, 117:708–722.

Zhang, K. & Nie, M. (2019). To pool or not to pool: Equilibrium, pricing and regulation. *SSRN Electronic Journal*.

Zhang, R. & Pavone, M. (2016). Control of robotic mobility-on-demand systems: A queueing-theoretical perspective. *The International Journal of Robotics Research*, 35(1):186–203.

Zhang, R., Rossi, F. & Pavone, M. (2019). Analysis, control, and evaluation of mobility-on-demand systems: A queueing-theoretical approach. *IEEE Transactions on Control of Network Systems*, 6(1):115–126.

Zhang, W., Wang, Q., Li, J. & Xu, C. (2020). Dynamic fleet management with rewriting deep reinforcement learning. *IEEE Access*, 8:143333–143341.

Zhou, Y., Li, Y., Hao, M. & Yamamoto, T. (2019). A system of shared autonomous vehicles combined with park-and-ride in residential areas. *Sustainability*, 11(11):3113.

Zhu, L., Garikapati, V., Chen, Y., Hou, Y., Aziz, H. M. A. & Young, S. (2018). Quantifying the mobility and energy benefits of automated mobility districts using microscopic traffic simulation. In *International Conference on Transportation and Development 2018*, pages 98–108.

Zhu, Z., Qin, X., Ke, J., Zheng, Z. & Yang, H. (2020). Analysis of multi-modal commute behavior with feeding and competing ridesplitting services. *Transportation Research Part A: Policy and Practice*, 132:713–727.

Zhu, Z., Xu, A., He, Q.-C. & Yang, H. (2021). Competition between the transportation network company and the government with subsidies to public transit riders. *Transportation Research Part E: Logistics and Transportation Review*, 152:102426.

Ziemke, D., Kaddoura, I. & Nagel, K. (2019). The MATSim open berlin scenario: A multimodal agent-based transport simulation scenario based on synthetic demand modeling and open data. *Procedia Computer Science*, 151:870–877.

Zitzler, E. (1999). *Evolutionary algorithms for multiobjective optimization: methods and applications*. Ph.D. thesis, ETH Zürich.