EPFL

Thèse n° 10 017

Modeling and Optimization of Ridesplitting Operations

Présentée le 25 juillet 2023

Faculté de l'environnement naturel, architectural et construit
Laboratoire de systèmes de transports urbains
Programme doctoral en génie civil et environnement

pour l'obtention du grade de Docteur ès Sciences

par

Caio Vitor BEOJONE

Acceptée sur proposition du jury

Prof. D. Licina, président du jury
Prof. N. Geroliminis, directeur de thèse
Prof. L. Leclercq, rapporteur
Prof. O. Cats, rapporteur
Prof. M. Bierlaire, rapporteur

École
polytechnique
fédérale
de Lausanne

2023

Behold, the present time, which alone we found could be called long, is abridged to the space scarcely of one day. But let us discuss even that, for there is not one day present as a whole. For it is made up of four-and-twenty hours of night and day, whereof the first has the rest future, the last has them past, but any one of the intervening has those before it past, those after it future. And that one hour passes away in fleeting particles. Whatever of it has flown away is past, whatever remains is future. If any portion of time be conceived which cannot now be divided into even the minutest particles of moments, this only is that which may be called present; which, however, flies so rapidly from future to past, that it cannot be extended by any delay. For if it be extended, it is divided into the past and future; but the present has no space. Where, therefore, is the time which we may call long? Is it nature? Indeed we do not say, "It is long," because it is not yet, so as to be long; but we say, "It will be long." When, then, will it be? For if even then, since as yet it is future, it will not be long, because what may be long is not as yet; but it shall be long, when from the future, which as yet is not, it shall already have begun to be, and will have become present, so that there could be that which may be long; then does the present time cry out in the words above that it cannot be long.

— St. Augustine, The Confessions (Book XI, paragraph 20)

To my parents and my beloved wife…

Acknowledgements

First, I would like to thank my supervisor Prof. Nikolas Geroliminis for giving me the opportunity to join LUTS, for providing an always friendly and inviting atmosphere, and for all the guidance, advise and encouragement that he offered me, that helped me succeed and grow as a person.
I would like to cordially thank all the members of my jury, Prof. Michel Bierlaire, Prof. Ludovic Leclercq, Prof. Oded Cats and Prof. Dusan Licina, for the time and energy they put in reviewing my thesis, as well as for all the constructive comments and the interesting discussion that we had in the oral exam.
I would like to thank our secretary Christine Debossens, for the immense kindness and support in every issue that came up, and her always joyful and inspiring presence. I am also grateful to all my colleagues in LUTS, older and newer, with whom we overlapped significantly and interacted the most during the last years.
There are no words to express my gratitude for my beloved wife Jéssica, who has been by my side in the past 11 year (!) supporting and encouraging me all the time. Not to mention the numerous times she took care of me when I was too tired (sometimes, I sure she was too, but still cared for me).
Last but not least, I am deeply grateful to my family, who have always been there for me and contributed the most in making me who I am today, who have been unconditionally supporting me and have encouraged me to pursue my goals, even when this meant that I needed to be far away from them.

Lausanne, 25/07/2023 Caio Vitor Beojone

Abstract

The advent of shared-economy and smartphones made on-demand transportation services possible, which created additional opportunities, but also more complexity to urban mobility. Companies that offer these services are called Transportation Network Companies (TNCs) due to their internet-based nature. Although ride-sourcing is the most notorious service TNCs provide, little is known about to what degree its operations can interfere in traffic conditions, while replacing other transportation modes, or when a large number of idle vehicles is cruising for passengers. Moreover, dynamic network-level models directly addressing ride-sourcing services can support the development of efficient strategies for both congestion alleviation and promotion of more sustainable mobility. Recent developments presented models focusing on ride-hailing (solo rides), but no work addressed ridesplitting (shared rides) in dynamic contexts. These models can be used for proper positioning of ride-sourcing drivers and improve vacant travel times, waiting times, and matching opportunities.

In Chapter 2, we experimentally analyze the efficiency of TNCs using taxi trip data from a Chinese megacity and an agent-based simulation with a trip-based MFD model for determining the speed. We investigate the effect of expanding fleet sizes for TNCs, passengers' inclination towards sharing rides, and strategies to alleviate urban congestion. We observe that, although a larger fleet size reduces waiting time, it also intensifies congestion, which, in turn, prolongs the total travel time. Such congestion effect is so significant that it is nearly insensitive to passengers' willingness to share and flexible supply. Our findings also show that, even if drivers quit the system in case of low profit, the system can converge in a fleet size, which still causes noticeable congestion.

In Chapter 3, we sought to develop a dynamic aggregated traffic network model capable of representing ride-sourcing services and background traffic in a macroscopic multi-region urban network. We combined the Macroscopic Fundamental Diagram (MFD) with detailed state-space and transition descriptions of background traffic and ride-sourcing vehicles in their activities to formulate mass conservation equations. We show that the model can accurately forecast the vehicles' conditions in near-future predictions (e.g., 30 minutes ahead). The development of this model prepares the path for developing real-time feedback-based management policies such as priority-based perimeter control or repositioning strategies for idle ride-sourcing vehicles and developing regulations over ride-sourcing in congested areas.

Chapter 4 evaluated the potential repositioning response of drivers when provided guidance

based on estimates of their earnings in a system offering ride-hailing (solo) and ridesplitting (shared) rides. The developed strategy provides a group of drivers with individualized near-future revenue estimates guiding drivers toward repositioning decisions that are more likely to maximize their earnings. Our main findings indicate that if the operator selects only a fraction of active drivers to provide guidance, these are likely to expect higher earnings than those without guidance. We also show that it manages to decrease the number of unserved requests compared to several state-of-art benchmarks at the same time that it increased vehicle occupancy and decreased the deadheading.
Chapter 5 presented a hierarchical control framework capable of repositioning vacant ride-hailing vehicles integrating model predictive control and coverage control in an urban traffic setting. The approach involved optimizing vehicle positions using near-future forecasts of the service and demand conditions, presenting a proactive strategy for dynamically deploying the fleet in advantageous spatial configurations. The proposed framework improved the performance in all tested scenarios by decreasing waiting times, increasing acceptance rates, and maximizing the usage of the available fleet.

Key words: transportation, ride-sourcing, human mobility, urban traffic, macroscopic fundamental diagram (MFD), simulation, vehicle repositioning, Markov chain, hierarchical control.

Résumé

L'arrivée de l'économie de partage et des smartphones a rendu possible les services de transport à la demande, ce qui a créé des opportunités additionnelles, mais a également amené encore plus de complexité à la mobilité urbaine. Nous nous référons aux entreprises qui fournissent ce type de service par les entreprises de réseau de transport (TNC) car elles sont naturellement basées sur Internet. Bien que le service de commande de véhicules avec conducteur soit le service le plus notoire que les TNCs présentent, il est toutefois peu évident à quel point son opération peut perturber les conditions de circulation lorsque ce service remplace d'autres moyens de transport ou lorsqu'un grand nombre de véhicules inactifs circulent à la recherche d'un passager. De plus, les modèles dynamiques à l'échelle du réseau traitant directement le service de commande de véhicules sont en mesure d'appuyer l'élaboration de stratégies efficaces tant pour réduire la congestion que pour promouvoir une mobilité plus durable. Les derniers développements ont présenté des modèles qui visent les services de commande de véhicules (trajets en solo), mais aucun autre travail n'a abordé les services de commande de véhicules avec un chemin partagé (trajets partagés) dans un contexte dynamique. Ces modèles peuvent servir de cadre pour positionner adéquatement les véhicules de commande et améliorer les temps de déplacement sans passagers, les temps d'attente, et les occasions de jumelage.

Dans le chapitre 2, nous testons l'efficacité des TNCs en utilisant les informations extraites de la base de données des trajets en taxi dans une mégapole chinoise, et une méthode de simulation multi-agents avec un modèle MFD basé sur les parcours pour déterminer la vitesse. Nous explorons l'effet de l'expansion de la taille des flottes de TNC, la volonté des passagers de partager leurs trajets, et les stratégies d'atténuation de la congestion urbaine. Nous constatons que, bien que les temps d'attente soient réduits pour une grande flotte, la flotte augmente la congestion, ce qui prolonge la durée totale du trajet. Un tel effet sur les embouteillages est si important qu'il devient presque insensible à la volonté des voyageurs de partager leurs parcours et à la flexibilité de l'offre. Nos résultats montrent également que même si les conducteurs abandonnent le système en cas de faibles revenus, le système peut converger dans une taille de flotte, ce qui continue de créer d'importants engorgements.

Dans le chapitre 3, nous avons cherché à élaborer un modèle dynamique de réseau de trafic agrégé capable de représenter les services de commande de véhicules avec conducteur ainsi que le reste du trafic dans un réseau urbain macroscopique multi-régional. Nous avons

rassemblé le diagramme fondamental macroscopique (MFD) avec une description élaborée de l'espace d'états et des transitions des véhicules TNC et le reste du trafic par activité. Cette étape est nécessaire pour formuler les équations de conservation de masse. Nous avons ensuite montré que ce modèle peut prédire les conditions du véhicule d'une façon très précise dans des prévisions à court terme (par exemple, 30 minutes à l'avance). Le développement de ce modèle permet de faire progresser les politiques de gestion en temps réel fondées sur la rétroaction, comme le contrôle du périmètre en fonction des priorités, ou les stratégies de repositionnement pour les véhicules TNCs inutilisés. Il permet également de développer des réglementations sur les services de contrôle des véhicules dans les zones encombrées.
Le chapitre 4 a évalué la réaction des conducteurs de véhicules à un avis de repositionnement possible fondé sur des estimations de leurs revenus dans un système proposant un trajet solo ou un trajet partagé. La stratégie développée fournit à un groupe de conducteurs des estimations personnalisées de rendu à court terme. Ainsi, elle oriente les conducteurs vers des décisions de repositionnement qui maximisent leurs revenus. Nos principales constatations indiquent que si l'opérateur ne sélectionne qu'une fraction de véhicules actifs pour fournir ces avis de repositionnement, ceux-ci sont susceptibles de s'attendre à des revenus plus élevés que ceux qui ne sont pas guidés. Nous montrons également que notre stratégie permet de réduire le nombre d'utilisateurs non desservis en comparaison avec d'autres stratégies de référence utilisant des technologies de pointe. Ceci est parce que notre stratégie en même temps augmente l'occupation des véhicules en service et réduit les déplacements haut-le-pied.
Le chapitre 5 a introduit un cadre de contrôle hiérarchique qui peut repositionner les véhicules non utilisés et vides. Ceci se fait en intégrant un modèle de contrôle prédictif et un contrôle de couverture dans le trafic urbain. L'approche a consisté à optimiser les positions des véhicules à l'aide de prévisions à court terme des conditions de la demande et du service. Ce cadre présente une stratégie proactive de déploiement dynamique de la flotte de véhicules dans des configurations spatiales avantageuses. Cette méthode proposée améliore le rendement dans l'ensemble des scénarios mis à l'essai en réduisant les temps d'attente, en augmentant les taux d'acceptation, et finalement en maximisant l'utilisation de la flotte de véhicules disponibles.
Mots-clefs : transport, service de commande de véhicules, mobilité humaine, trafic urbain, diagramme fondamental macroscopique (MFD), simulation, repositionnement des véhicules, chaîne de Markov, contrôle hiérarchique.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation and background

One of the most prominent innovations seen throughout streets around the world is the ubiquitous presence of drivers using their vehicles for on-demand transportation services. Companies use mobile applications connected through the internet to match these drivers and their passengers in real-time. Due to the nature of their operations, these companies are called Transportation Network Companies (TNCs), but the service itself is called ride-sourcing, e-hailing, and ride-sharing, for instance Rayle et al., 2016. Ride-sourcing services have revolutionized mobility concepts for on-demand transportation as a result of the advantages they provide, such as convenience, door-to-door rides, low fares, etc. On-demand transportation services sound as a promising direction to improve mobility and fight car ownership. Moreover, many TNCs offer, among the service options, shared rides (called ridesplitting). These services try to match passengers with a reasonably similar trip within a time window. For TNCs and drivers, this service may yield increased profits if it is capable of matching passengers and drivers efficiently. For the passengers, this service presents a cheaper option, but they might face longer travel distances/times. Passengers may also have almost the same advantages as those from a taxi service as door-to-door rides and no need to search for parking. In general, these services seem to have a positive impact on economic efficiency (S. T. Jin et al., 2018; Tachet et al., 2017).

Naturally, all this expansion raised several concerns regarding TNCs' operations. Oppositely to taxis, TNCs face no limitation, in most cities, on the fleet size that can operate, no price control, service requirements, and other legal obligations faced by the taxi industry. Moreover, as these services base their operations on mobile applications connected to the internet, there are concerns over issues of data privacy and security (S. T. Jin et al., 2018). Rogers, 2017 adds other social costs, such as diminished safety and lack of professional training. Proper planning and regulation have vital importance in the development of shared transportation for the near future (Narayanan et al., 2020). Another point of concern is the surge pricing models used, which may considerably increase the fares in moments that driver availability is insufficient

(Schwieterman & Smith, 2018). On the other hand, surge pricing mitigates the potential chaos of a bargaining process. Notably, it handles the spatial-temporal imbalances between driver supply and rides demand (Dong et al., 2018). Moreover, recent results point out that it can increase drivers' revenues but make customers worse during highly surged periods (Zha et al., 2018).

Although it is not clear whether ride-sourcing is beneficial or unfavorable (or whether it causes anything significantly) for traffic congestion, the path to clear it is to understand how it is replacing traditional transportation modes. In case ride-sourcing trips are directly substituting private vehicles or taxis trips then, they should have a secondary influence on congestion (Erhardt et al., 2019). However, if ride-sourcing competes with public transportation modes (buses, trains, metro) or inducing latent demand, then the effects on congestion should be significant. Additionally, it might increase vehicle kilometers traveled (VKT) when vehicles cruise for passengers or when it induces latent demand (Vinayak et al., 2018). A probable scenario for Tirachini and del Río, 2019 and Tirachini and Gomez-Lobo, 2019 has ride-sourcing extensively substituting transit but only inducing latent demand to a small extent. In a recent survey across TNC users in San Francisco (Rayle et al., 2016), in a question "How would you have made this trip if TNC service was not available?", 40% answered by taxi, 33% by bus, and only 6% by car. Thus, TNC can be an attractive alternative for public transport users, and, combined with a large number of empty vehicles, it can create additional congestion problems. The consequences of such non-cooperative interactions can be catastrophic for urban traffic (Çolak et al., 2016; Olmos et al., 2018; Roughgarden, 2005). For instance, reductions in demand for buses can cause imbalances making them miss their schedule, dropping their capacity because of bus bunching (Saw et al., 2019; Sirmatel & Geroliminis, 2018b). The transportation literature observes these effects for decades (W. S. Vickrey, 1969).

1.2 Challenges

1.2.1 Ride-sourcing with ridesplitting and traffic congestion

It is imperative to understand how TNCs' operations can interfere in traffic conditions while replacing other transportation modes to seek improvements in urban mobility. Foremost, this understanding must cover the performance of traffic and operations. It is critical to relate the fleet size with the average speeds and service level, which are related to mobility and accessibility measures (Hanson & Giuliano, 2017; Páez et al., 2012). The literature already presents evidence of traffic improvements from the use of curbside parking for idle drivers (Xu et al., 2017). Hall et al., 2018 shows that ride-sourcing can complement public transport activities, and speculates that users avoiding the limitations of fixed-route and schedule modes are the reason behind the complementary effect. Moreover, the matching process shall have a place, and thus the impact of passengers' behavior too, in a ridesplitting scenario. For this reason, Wei et al., 2020 used logit models to detail the decisions of drivers and passengers in a multi-modal setting and showed that ride-sourcing has the potential to decrease traffic

performance. Much of the literature on ride-sourcing relies on surveys (Alemi et al., 2018; Dong et al., 2018; Lavieri & Bhat, 2019; Rayle et al., 2016; Vinayak et al., 2018), economics (He & Shen, 2015; Xu et al., 2017; Zha et al., 2018; Zha et al., 2016), and data regressions (Contreras & Paz, 2018). These studies became available because of the availability of large datasets on human mobility, which enabled studies not only for ride-sourcing but for all transportation modes, such as buses (Bassolas et al., 2020) and taxis (Hamedmoghadam et al., 2019; Riascos & Mateos, 2020). Even though some surveys, such as Wenzel et al., 2019, Tirachini and Gomez-Lobo, 2019, Zha et al., 2016, link ride-sourcing services with increased traffic, they do not consider the dynamics of congestion directly nor how these services affect urban mobility and influence congestion. Nourinejad and Ramezani, 2020 applies pricing strategies in a dynamic non-equilibrium model that tracks riders and drivers and the respective market performance measures.

However, one must acknowledge the research efforts towards modeling the dispatch of taxis and shared taxis operation. Lee et al., 2004 improved the dispatch of taxis using actual travel distance instead of Euclidean distances to passengers. Wong and Bell, 2006 added traffic congestion to the dispatch process. More recently, Ramezani and Nourinejad, 2018 used a macroscopic model to control taxi fleets in a multi-region setting. Martinez et al., 2015 used an agent-based simulation to show the potential of shared-taxis for improving mobility management in urban areas. Santi et al., 2014 developed shareability networks to enable the operation of shared-taxis in New York City. Hosni et al., 2014 presented a formulation for the problem of assigning passengers to taxis and computing the optimal routes of taxis. Jung et al., 2016 used hybrid-simulated annealing for dynamic shared-taxi dispatch. Research efforts also focused on ride-sharing systems. Alonso-Mora et al., 2017 further developed the use of shareability networks to allow real-time dispatch in ride-sharing systems. Stiglic et al., 2016 assessed the impacts of riders' and drivers' flexibility to foster the use of ride-sharing. Nourinejad and Roorda, 2016 showed that a decentralized approach for ride-sharing could have higher user cost savings and vehicle kilometers traveled (VKT) savings. Vazifeh et al., 2018 presented a real-time minimum fleet problem for on-demand urban mobility using New York City taxi data. In Long et al., 2018, the authors address the problem of travel time uncertainty in ride-sharing services. Furuhata et al., 2013 and Agatz et al., 2012 present insightful literature reviews on ride-sharing services. Finally, other researchers worked on pick-up and delivery problems, and, more specifically, dial-a-ride problems. Cortés et al., 2010 formulated a pick-up and delivery problem with transfers. Berbeglia et al., 2010 reviewed dynamic pick-up and delivery problems, as dial-a-ride problems. Masmoudi et al., 2018 presented a dial-a-ride problem with battery swapping. Bongiovanni et al., 2019 proposed a variant of dial-a-ride problems for electric-autonomous vehicles. Molenbruch et al., 2017 and Ho et al., 2018 reviewed dial-a-ride problems, their solution methods, and classifications. Nonetheless, these problems do not correlate TNCs' fleet size to traffic conditions, nor the passengers' behavior yet. Ignoring the effect of congestion in the operation of ride-sourcing and ridesplitting services can influence the conclusions made. Alonso-Mora et al., 2017 showed that it is possible to serve the taxi demand of Manhattan, with reductions of 30% on

the current fleet. The paper assumes that all passengers are willing to share a ride with others and that the system has perfect information about future demand. They also did not consider the effect of congestion due to different demand conditions or the compliance of the taxi companies to decrease their fleet size.

It is worth mentioning that other studies contributed to the understanding of labor supply related to surge pricing (Zha et al., 2018) and multi-modal traveler decision making (Su & Wang, 2019; Wei et al., 2020). However, given the complexity of real-time matching algorithms and the dynamic nature of traffic, they usually avoid a spatial representation of the urban network and the ridesplitting matching process. Instead, they rely on equilibrium models.

1.2.2 Forecasting ride-sourcing and traffic conditions

Ride-sourcing operators might not have a direct interest in congestion (see, for example, Beojone and Geroliminis, 2021b), but a dynamic model that captures congestion can be valuable for various operational decisions. Efficient repositioning depends not only on demand knowledge but also on the dynamics of idle vehicles and the time needed to move from one region to the other. This requires a proper dynamic representation of various states, and we can expand the MFD concept of aggregated modeling in this direction. Then, if an MFD model could capture critical features of ride-hailing and ridesplitting services, one could investigate various management schemes.

Since the first effort on conceptualizing MFD-based perimeter control (Daganzo, 2007), similarly to the approaches of gating, appearing in ramp-metering and signal control, multiple control schemes have been developed. They include various control methods for heterogeneously congested cities partitioned in homogeneous regions, such as Model Predictive Control (MPC), Proportional-Integral control, optimal control, etc (Haddad & Shraiber, 2014; Kouvelas et al., 2017; Sirmatel & Geroliminis, 2018a). Models based on the MFD dynamics can forecast near-future conditions of urban systems with lower computational run times than costly simulations, avoiding demanding and detailed route choice and assignment frameworks. They also require a small number of input data that is more realistic to obtain. Ramezani and Nourinejad, 2018 presented the first effort for taxi repositioning control using MFD models, and it had a similar formulation to Geroliminis, 2015, which studied cruising-for-parking.

In most MFD control papers, accumulation-based MFD models are utilized. These models require the existence of an Outflow-MFD, which considers a unimodal low scatter relation between trip completion and accumulation, assuming a steady-state relationship between production (veh.km travelled per unit time in a region) and outflow that implies a memoryless constant average trip length. Such an assumption might be problematic for a dynamic ride-sourcing model as the memoryless trip length assumption will not hold. On the other side, trip-based models (which consider the existence of a Speed-MFD and track the remaining trip length distribution) are computationally demanding and problematic to integrate with control. A newly developed MFD model, named M-model, provides a decent approximation

of trip-based models (Lamotte et al., 2018; Murashkin, 2021; Sirmatel & Geroliminis, 2021). If we can formulate a ride-sourcing model including M-model traffic dynamics, it can be a powerful tool for network-level control, e.g., for relocating vacant vehicles when demand is not balanced, shifting from reactive strategies, where a passenger first leaves unserved so the operator acts to avoid further losses.

Nevertheless, multiple challenges arise when modeling ride-sourcing services with a macroscopic traffic framework. Firstly, there is a more complex state representation and interactions due to different activities, for examples cruising with no passengers, ride-hailing passengers or ridesplitting passengers, etc. Namely, the developed framework should be able to provide an understanding of how drivers transition from one activity to another, how these transitions occur in a multi-regional setting, and how they interact with other elements in the traffic system (e.g., passengers and background traffic). Including a ridesplitting service option not only adds more activities but also adds different dynamics for dealing with passenger-driver matching and how it affects the main movements of drivers. Especially, serving on a *first-come-first-served* (FCFS) basis requires the understanding that an incoming request can interrupt an ongoing service to assign a new passenger to a shared ride. Furthermore, passengers have different pick-up and drop-off locations, possibly, all in distinct regions, requiring a driver to deliver the passengers far from each other. Besides the movement of passengers and drivers in the traffic system, ridesplitting services present significant market thickness in that increases in the demand creates a positive feedback in the service capability of serving multiple passengers. It all challenges the representation assignments and losses integrated in the same model framework, since macroscopic models do not track individual trips. Overcoming such challenges is imperative in the described context, where near-future forecasts are essential for developing various managerial frameworks.

1.2.3 Improving ride-sourcing for customers and drivers through repositioning

In a daily basis, geographical variations on the demand can create an imbalance between the ride-sourcing service demand and supply of drivers to serve it. It is of the best interest of the Transportation Network Company (TNC) responsible for the service operation to balance both demand and supply to maintain a satisfactory service quality. Among the actions the TNC can take, one can include trying to attract more drivers to the service, or using surge-pricing, or a combination of these to take drivers to undersupplied areas. However, the fleets of this service are formed by human drivers that offer rides to make profits and are free to make a series of decisions. They include defining where they will look for new assignments and when to offer rides.

Therefore, the TNC is not able to deploy extra vehicles whenever there is a shortage, and it requires convincing ability to relocate the available pool of drivers. Lu et al., 2018 and Sadeghi and Smith, 2019, and Powell et al., 2011 incentivized drivers to decide on the best location for the next assignment through mechanisms that control the supply of drivers and

the demand of passengers, such as surge pricing (Castillo et al., 2018). However, the issue with these strategies is their reactive nature, in the sense that they mainly accounted for past events (i.e., lost demand) or current conditions. For instance, if an area faces recurrent losses of requests, customers will likely change their travel option to a more reliable transportation mode. From the drivers' perspective, they face high uncertainty whether an area previously marked as 'high demand' or 'surge price multiplier' will be crowded with vacant vehicles, resulting in longer cruising times. Such problem also emerges because these strategies only provide limited information for drivers (in time and scope), which must estimate themselves the most profitable option.

With a more proactive nature, the literature has explored the use of Markov Decision Processes (MDP) to understand drivers and their decision-making in the search for next passengers (Shou et al., 2020). For instance, X. Zhou et al., 2020 proposed a MDP that recommends cruising directions to taxi drivers. Yu et al., 2019 formulated a Markov Decision Process (MDP) so that profit maximization objectives can be taken into account. However, these approaches may fall short in assuming that individual drivers understand the dynamics and are able to compute themselves the rewards of their decisions. Moreover, they ignore the impact of competition between individual drivers, while still assuming drivers compliance to provided guidance.

Other strategies emerge from the optimization of passenger-driver matching algorithms. These strategies include Alonso-Mora et al., 2017, Wallar et al., 2018, Simonetto et al., 2019, and Liu and Samaranayake, 2020 who sent empty vehicles to the location of recently unsatisfied customers, suffering from the shortcomings of their reactive nature. Other examples can be found in Afeche et al., 2018, Yu et al., 2019, and Wang and Yang, 2019 and references therein. More recent studies, try to take actions before losing these passengers. Zhu et al., 2022 uses coverage control to proactively position idle drivers in areas more likely to originate new requests. Ramezani and Nourinejad, 2018 uses an MPC to relocate idle taxis in a macroscopic set of regions. However, although highly optimized and proactive, these approaches assume complete compliance to the provided instructions. Hence, they ignore that the objectives of individual humans offering rides, or they assume that the TNC owns a completely autonomous fleet.

From all the above, multiple challenges arise when timely balancing demand and supply of ride-sourcing by repositioning the currently available drivers. The first challenge is to take the burden of identifying the most profitable options from drivers with limited information, which only have access to limited information while accouting for their future activities. However, these options must account for a driver's future activities and their interactions. For instance, Beojone and Geroliminis, 2023a depicted the macroscopic interactions of ride-sourcing activities in a multi-regional traffic system. Nevertheless, it still requires the translation to the individual level, which cannot be approximated in a "continuum" of drivers (represented by real numbers, instead of integers). Thus, translating revenues is no longer a process of counting and summing discrete completed ride requests. Along with the revenue estimation,

a challenge remains in giving positions that minimizes unnecessary demand coverage overlapping among drivers. Finally, if the goal is to convince drivers to reposition, the strategy must ensure that compliant drivers have an improved outcome from reliable predictions.

1.2.4 Autonomous vehicles for repositioning ride-sourcing vehicles

As mentioned earlier, spatiotemporal variations in demand can create supply imbalances between drivers and passengers, manifesting as deterioration of system efficiency and service quality. Additionally, since the service rely on human drivers, there is the additional challenge of persuading them. However, the advent of autonomous mobility has driven attention and TNCs are among the leaders in developing and deploying such an innovation (Ohnemus & Perl, 2016). Therefore, one can envision the implementation of a completely autonomous fleet operated and controlled by the TNC. With such a scenario, the TNC has the opportunity of coordinating every repositioning decision simultaneously and continuously in all operating vehicles.

Focusing on repositioning with autonomous vehicles (deliberately or inferred), the literature has examples from different natures.[I] In Alonso-Mora et al., 2017 and Simonetto et al., 2019, the passenger-matching strategy was extended to reposition idle vehicles by assigning them to the location of recently unserved requests, coordinating their movements to minimize the distance traveled. Wallar et al., 2018, in the other hand, proposed an algorithm to partition the fleet into rebalancing regions, determine a real-time demand estimates and an algorithm to assignment of idle vehicles to these rebalancing regions. Liu and Samaranayake, 2020 used a probabilistic rebalancing method based on demand distributions obtained from historical data to guide idle vehicles to areas with a high probability of future requests. Finally, recall Zhu et al., 2022, which coordinated the movement of all idle drivers using coverage control. One issue with the aforementioned works is that the operators only take past events (e.g., unserved/lost requests) or the current situation into account, due to their reactive nature.

However, in the application of vehicle repositioning, forecast of future conditions is crucial for deploying the fleet proactively towards improved performance. The complexity of the operations due to large fleet size, demand uncertainty and spatiotemporal heterogeneity in the distribution of congestion make the control problem a challenging one. With a different perspective from the other studies, Ramezani and Nourinejad, 2018 used macroscopic modeling and control to reposition taxi vehicles in a multi-region setting by minimizing the delay of drivers and passengers. Ramezani and Valadkhani, 2023 and Valadkhani and Ramezani, 2023 extends the MFD-based model from Ramezani and Nourinejad, 2018 while formulating matching and repositioning control algorithms for ride-hailing. The idea is based on the concept that, given an urban network partitioned into several homogenously congested regions, models based on the macroscopic fundamental diagram (MFD) can describe the dynamics of

[I] Note that we infer that vehicles are autonomous when the individual human driver decision process is not evaluated, implying a complete compliance with the repositioning instructions.

traffic states while requiring no information on the exact condition of each individual vehicle.

Besides repositioning, the MFD concept is behind different control methods built on the known concept of gating, appearing in ramp metering or signal control implementations in cities for traffic management can control. A few examples are: Multi-region Proportional–integral control (Aboudolas & Geroliminis, 2013; Ding et al., 2018; Ingole et al., 2020), optimal control (Aalipour et al., 2019; Haddad, 2017), robust control (Ampountolas et al., 2017; Haddad, 2015; Haddad & Shraiber, 2014; Y. Li et al., 2021; Mohajerpoor et al., 2020; Zhong et al., 2018), adaptative control (Haddad & Mirkin, 2016; Haddad & Zheng, 2020; Kouvelas et al., 2017), control with route choice (Menelaou et al., 2017; Menelaou et al., 2019), hierarchical control (Fu et al., 2017; Ramezani et al., 2015), and even low-altitude air traffic control (Haddad et al., 2021; Safadi et al., 2023). Model predictive control methods have also been combined with MFD dynamic models, for example Model Predictive Control (MPC) with MFD-based travel time and delays (Csikós et al., 2017), hierarchical MPC with MFD-based and link-level models (Z. Zhou et al., 2017), MPC with perimeter control and regional route guidance (Sirmatel & Geroliminis, 2018a), and extensions with a path assignment mechanism (Yildirimoglu et al., 2018), demand and state estimation (Kumarage et al., 2023), multi-scale stochastic MPC with connected vehicles (K. Yang et al., 2018), or combined operation of state estimation and MPC (Sirmatel & Geroliminis, 2020). Then, various control methods focused on MFD-based modeling have been developed for heterogeneously congested cities partitioned in a number of homogeneous regions (Haddad & Shraiber, 2014; Kouvelas et al., 2017; Sirmatel & Geroliminis, 2018a). In Sirmatel et al., 2021, past traveling behavior is summarized into a total remaining distance state and this information is used to modify the exit function of multi-region MFD-based models, capturing thus the effect of trip length variability which is relevant e.g. in control applications. However, most of these works are based on the regional scale, therefore they can only give an aggregated command, i.e., the transfer flow from one region to another, but still lack a complete structure to link the repositioning strategy from a macroscopic decision to the detailed instructions for individual vehicles.

Herein, the adoption of an autonomous fleet enables the decision process to focus solely on system-wide performance, instead of individual driver objectives. In the other hand, different challenges emerge for repositioning autonomous vehicles. The first challenge is to coordinate the movements of all vehicles. For instance, it should avoid instructions causing coverage overlap decreasing the system capacity of serving incoming requests. In large-scale environments with fleets of hundreds or thousands of vehicles, the complexity of any proposed solution can quickly turn the problem unsolvable or, at least, infeasible for real-time solutions. The idea of breaking the repositioning problem in different layers, with different scales and scopes each, is a plausible path. However, it raises the concern of balancing the different layers such that the envisioned gains in computational complexity are obtained.

1.3 Objectives

The first and foremost objective in this thesis is to contribute to the society and existing research by improving our understanding about ride-sourcing services, its interactions in the urban area, and improve it to those closely related to it (drivers and passengers) while keeping on perspective the surrounding sustainability dimensions. Based on the preceding literature review on the main thematic areas of the thesis and the personal interest of the author, there exist a set of research areas that can be further explored. Therefore, the following objectives are set, organized in chapters following the structure of the thesis.

- **Chapter 2: On the inefficiency of ride-sourcing services towards urban congestion.** This chapter aims to investigate the effect of expanding fleet sizes for TNCs, passengers with different willingness to share, and operational strategies over congestion conditions under a sustainable perspective. The investigation considered a trip-based MFD traffic model integrated into an event-based simulation to tackle the dynamics of congestion. The traffic model considers private vehicles and TNCs' vehicles. The dynamics of the system are based on an aggregated dynamic traffic model, the network Macroscopic Fundamental Diagram (MFD) (Geroliminis & Daganzo, 2008; Loder et al., 2019), to avoid the computational burden of micro-simulation and the lack of sufficient data for proper calibration. We model interactions between travelers and vehicles with an efficient matching algorithm. It is beyond the scope of the chapter, the mode-choice modeling. Hence, we focus on the supply of rides and its participation in traffic dynamics testing several fleet sizes and willingness to share to cover a wide range of scenarios with various values of these critical variables defined externally.

- **Chapter 3: A dynamic multi-region MFD model for ride-sourcing with ridesplitting.** We develop an MFD-based model representing ride-sourcing services and background traffic in a macroscopic multi-region urban network. The modeled ride-sourcing service offers ride-hailing (single rides) and ridesplitting (shared rides). Model states describe drivers on their ongoing activities and regions. We evaluated the proposed model by comparing the errors of the proposed model with benchmarks from the literature utilizing a detailed agent/trip-based simulator developed with real data from the the central business district of Shenzhen, China. Additionally, a sensitivity analysis investigated the generability of the model to assess the performance of multi-region traffic systems to several service parameters, such as fleet size, willingness to share or waiting time tolerance.

- **Chapter 4: Guiding the relocation of ride-sourcing drivers with revenue forecasting.** We evaluate the potential repositioning response of drivers when provided an estimate of their earnings. Drivers are not forced to specific actions, but may not comply with the provided guidance. They base their decisions on their earning expectations for remaining idle in the current region or repositioning to a neighboring one. The operator uses a mixed discrete-continuous time Markov chain (MDCTMC) to estimate individual

earnings for a given decision in the short-term. A microscopic process identifies the positions and their associated paths that provide the highest chances of matching. Finally, the performance of the proposed strategy is compared to state-of-art strategies from the literature, such as Coverage control (Zhu et al., 2022) and dispatching near drivers to areas of recently lost requests (Alonso-Mora et al., 2017).

- **Chapter 5: A hierarchical control framework for vehicle repositioning in ride-hailing.** In this chapter, we propose a hierarchical control strategy for the relocation of idle ride-sourcing vehicles, for addressing the gap between proactive repositioning strategies and micro-management of vehicles in such activities. The upper-layer utilizes an aggregated model, which is an approximation of trip-based MFD modeling approach (building on Beojone and Geroliminis, 2023a). A model predictive control (MPC) framework is employed to determine the number of idle vehicles to be relocated for each pair of regions. Unlike perimeter control MPC methods, fleet management MPC requires the integration of more sophisticated MFD-based models describing mixed dynamics of private vehicles and taxis. In the lower-layer, given the demand density over the current region, a coverage control scheme operates to distribute the vehicles within the region to achieve a demand-aligned configuration, which provides each vehicle with relatively detailed (i.e., intersection/node-level) position guidance. To bridge both layers, a middle-layer mechanism is developed for converting the upper-layer decisions into dispatching commands for individual vehicles by solving an Assignment Problem, which minimizes the distance required to achieve the optimal coverage and repositioning decisions. An agent-based simulator built on a trip-based MFD model is utilized with empirical taxi data from a real network of Shenzhen for validation of the proposed strategy. Note that, differently from the previous Chapter 4, the intent is to identify the most optimized repositioning strategy, assuming vehicles full compliance with the provided instructions (i.e., autonomous vehicles operated by the service provider), creating a best case scenario for repositioning decision benchmarks.

1.4 Contributions

Driven by the stated objective and based on the methods and results that will be elaborated in detail in the next chapters, the research conducted in the scope of this thesis leads to the following contributions, listed and elaborated per chapter as follows.

- **Chapter 2: On the inefficiency of ride-sourcing services towards urban congestion.** Our findings show that ride-sourcing can lose lose attractiveness to public transport when fleets are large enough to cause negative traffic externalities. Nevertheless, a higher willingness to share can minimize waiting and travel times. In summary, the contributions are the following: i) To the best of our knowledge, this is among the first works to quantitatively relate the TNCs' fleet size, willingness to share with a dynamic traffic congestion model; ii) It investigates the effect of the previously mentioned features and

'empty' vehicles (number of cruising vehicles without passengers) on the performance of the system; iii) It develops a parking management policy for cruising vehicles that can mitigate negative congestion externalities while maintaining the same quality of service; and iv) It shows that, if drivers adapt their participation in a day-to-day basis due to low profit for certain wage thresholds, the number of drivers can be above the ideal for traffic and service quality.

- **Chapter 3: A dynamic multi-region MFD model for ride-sourcing with ridesplitting.** To the best of our knowledge, this is among the first attempts to present and evaluate such a model for ride-sourcing services with the option for ridesplitting (shared rides) in the literature, providing a likely path to the development of strategic repositioning and regulatory problems on ride-sourcing services. It provides an structured state space, such that the size of the model grows according to a second-order polynomial as a function of the number of regions and linearly as a function of the passenger-capacity of ridesplitting trips. It also depicts the dynamics associated with ridesplitting activities, which incur in a unforeseen state transition in previous MFD-based models where a trip is interrupted before completion. We use the developed model to provide insights on the service performance, comparable to detailed simulations but with much shorter computational times. Finally, the chapter presents a benchmark comparison, where the proposed model had superior accuracy than other models in the literature.

- **Chapter 4: Guiding the relocation of ride-sourcing drivers with revenue forecasting.** In a simulated study, based in Shenzhen, China, we compare the performance of guided drivers and unguided ones. The proposed method has a flexible structure partitioning the guidance in the identification of the best repositioning decision and the estimation of drivers' revenues. First, an elegant approximation identifies the areas where a driver will have the highest chances of getting matched. Second, a Markov Chain model depicts a driver's activities associated with a potential repositioning path and translate them into revenue forecasts. We show that guided drivers have increased revenues and are likely to follow the provided guidance in the long term. We also show that it is possible to minimize the unserved requests without providing guidance to every driver. The proposed method achieved superior results compared to the reactive approaches, in terms of lost requests, vehicle occupancy and deadheading. Most of the results for guided drivers were comparable to those of the coverage control, but without assuming drivers complete compliance with the provided guidance. The proposed method has also little sensitivity to drivers' decision process.

- **Chapter 5: A hierarchical control framework for vehicle repositioning in ride-hailing.** Compared with control strategies using exclusively the upper- or the lower-layers, the results indicate that the proposed hierarchical framework method yields performance improvements by answering more requests with lower waiting times. Therefore, it presents a new best case scenario for benchmark evaluation of repositioning strategies. Additionally, it presents a general and flexible structured strategy that enables modifica-

tions in each layer for testing faster algorithms and solutions for repositioning of idle vehicles.

1.5 Structure

The thesis is organized in 5 chapters, separated by thematic area. The interior structure of each chapter is described and the respective publications of parts of each chapter in conferences and in scientific journals are listed below. Chapters 2, 3, 4, and 5 are standalone articles published or under review in scientific journals. The introductions of these articles were largely summarized to minimize the overlapping with the content presented in Chapter 1. Each chapter has its own independent notation, i.e., the same symbol can be used to represent different quantities in different chapters.

Chapter 2 evaluates the influence of ride-sourcing services in the urban environment using discrete-event simulations. The simulation framework is explained among its features, including the entities in the environment, their relationships and their interactions with traffic conditions and public transportation options. The service performance is evaluated alongside traffic measurements and the outcomes for involved stakeholders (including passengers, drivers, service operator). A strategy based on off-street parking is presented as a form to mitigate potential externalities of ride-sourcing services. Parts of this research were presented in:

- Beojone, C. V., & Geroliminis, N. (2019). Accessing shared mobility impacts on urban traffic networks through discrete event simulations. *8th Symposium of the European Association for Research in Transportation (hEART 2019)*
- Beojone, C. V., & Geroliminis, N. (2020b). Towards sustainable ride-sourcing services: a simulation study on the effects of congestion, fleet size and willingness to share. *99th Transportation Research Board Annual Meeting (TRB 2020)*
- Beojone, C. V., & Geroliminis, N. (2021c). A path to take passengers from single to shared rides: a study on ridesplitting. *9th Symposium of the European Association for Research in Transportation (hEART 2020)*
- Beojone, C. V., & Geroliminis, N. (2021b). On the inefficiency of ride-sourcing services towards urban congestion. *Transportation Research Part C: Emerging Technologies*, *124*, 102890. https://doi.org/10.1016/j.trc.2020.102890 **(Chapter 2 is a version of this publication)**

Chapter 3 presents an MFD-based model to evaluate and predict near-future conditions of ride-sourcing service including traffic conditions and drivers activities. The model is built without assuming steady-state conditions and memoryless trip lengths. Drivers' intra- and inter-regional movements in ride-sourcing activities are detailed along with the background traffic.

The model is evaluated in a multi-region sensitivity analysis and compared to benchmarks in a two-region case.

- Beojone, C. V., & Geroliminis, N. (2020a). Macroscopic modeling of ride-sourcing services. *20th Swiss Transport Research Conference (STRC 2020)*
- Beojone, C. V., & Geroliminis, N. (2021a). Macroscopic ride-sourcing model for fleet rebalancing. *100th Transportation Research Board Annual Meeting (TRB 2021)*
- Beojone, C. V., & Geroliminis, N. (2022a). A dynamic multi-region mfd model for ride-sourcing systems with ridesplitting. *101st Transportation Research Board Annual Meeting (TRB2022)*
- Beojone, C. V., & Geroliminis, N. (2023a). A dynamic multi-region mfd model for ride-sourcing systems with ridesplitting. *Transportation Research Part B: Methodological, (Under Review)* **(Chapter 3 is a version of this publication)**

Chapter 4 presents a strategy to persuade a group of ride-sourcing drivers to reposition such that their revenues are maximized and service quality is improved. It presents a process that identify optimized destinations for repositioning decisions such that the expected number of matchable requests are maximized. A mixed discrete-continuous time Markov chain (MD-CTMC) is developed to provide individualized forecasts of near-future activities, while drivers are assumed to decide about repositioning through a logit process. The proposed strategy is evaluated regarding its ability to retain drivers' long term compliance. Additionally, the proposed strategy is compared to a series of benchmarks from the literature.

- Beojone, C. V., & Geroliminis, N. (2021d). Repositioning idle vehicles in ridesplitting operations using pricing. *21st Swiss Transport Research Conference (STRC 2021)*
- Beojone, C. V., & Geroliminis, N. (2022b). An optimized driver repositioning strategy in ridesplitting with earning estimates: a two-layer dynamic model and control. *10th Symposium of the European Association for Research in Transportation (hEART 2022)*
- Beojone, C. V., Geroliminis, N., & Yin, Y. (2022). Repositioning ridesplitting vehicles through pricing: a two-region simulated study. *11th Triennial Symposium on Transportation Analysis (XI TRISTAN)*
- Beojone, C. V., & Geroliminis, N. (2023c). Relocation incentives for ride-sourcing drivers with path-oriented revenue forecasting based on a markov chain model. *Transportation Research Part C: Emerging Technologies, (Under Review)* **(Chapter 4 is a version of this publication)**

Chapter 5 presents a multi-layer control framework to reposition ride-sourcing vehicles that minimizes the number of lost requests. It briefly presents the three layers and their interactions.

The upper-layer is formed by a MPC controller deciding the number of vehicles repositioning. The middle-layer performs an assignment optimization to minimize the distance drivers will have to travel to fulfil the instructions from the other layers. Since this is a collaborative effort, the focus of this chapter is the middle-layer, where the candidate's contributions are more pronounced. It must be highlighted also that this is not a complete study in this version of the thesis. The current version of this study is an extended abstract accepted for a full paper submission in ISTTT25 (25$^{\text{th}}$ International Symposium on Transportation and Traffic Theory).

- Beojone, C. V., Zhu, P., Sırmatel, I. İ., & Geroliminis, N. (2023). A hierarchical control framework for vehicle repositioning in ride-hailing systems. *25th International Symposium on Transportation and Traffic Theory (accepted for full paper submission)* **(Chapter 5 is a version of this manuscript)**

Finally, Chapter 6 summarizes the findings and contributions of the thesis while discussing about interesting future research directions in the field.

2 On the inefficiency of ride-sourcing services towards urban congestion

This chapter is based on the paper:

- Beojone, C. V., & Geroliminis, N. (2021b). On the inefficiency of ride-sourcing services towards urban congestion. *Transportation Research Part C: Emerging Technologies, 124*, 102890. https://doi.org/10.1016/j.trc.2020.102890

2.1 Introduction

It is imperative to understand how TNCs' operations can interfere in traffic conditions (favorable or unfavorable, if any effect at all) while replacing other transportation modes to seek improvements in urban mobility. Simultaneously, it is critical to relate the fleet size with the average speeds and service level, which are related to mobility and accessibility, influencing passengers and social welfare. Ignoring the effect of congestion in the operation of ride-sourcing and ridesplitting services can influence the conclusions made in any work as much as simplifying the service operations.

Therefore, this chapter aims to investigate the effect of expanding fleet sizes for TNCs, passengers with different willingness to share, and operational strategies over congestion conditions under a sustainable perspective. The investigation considered a trip-based MFD traffic model integrated into an event-based simulation to tackle the dynamics of congestion. The traffic model considers private vehicles and TNCs' vehicles. The dynamics of the system are based on an aggregated dynamic traffic model, the network Macroscopic Fundamental Diagram (MFD) (Geroliminis & Daganzo, 2008; Loder et al., 2019), to avoid the computational burden of micro-simulation and the lack of sufficient data for proper calibration. We model interactions between travelers and vehicles with an efficient matching algorithm. It is beyond the scope of the paper, the mode-choice modeling. Hence, we focus on the supply of rides and its participation in traffic dynamics testing several fleet sizes and willingness to share to cover a wide range of scenarios with various values of these critical variables defined externally.

Following the motivation and the challenges raised about ride-sourcing with ridesplitting and traffic congestion, which is given in Section 1.2.1 of Chapter 1, the remainder of this chapter is organized as follows. Section 2.2 describes the methodological framework, including the simulator architecture, the real data, the matching process for passengers, and a parking-oriented strategy to decrease the circulation of empty vehicles. Then, Section 2.3 presents numerical results on the effect of fleet size, willingness to share, and parking policies in the quality of service and on network congestion.

2.2 Data and Methodology

2.2.1 Data description

The original data contain GPS coordinates of 199'819 trips, with their respective origins and destinations, of 20'000 taxis every 30 seconds for 20 hours in the city of Shenzhen, China. Shenzhen is immediately north of Hong Kong, in the southern province of Guangdong. Due to a rapid growth period, the population was close to 11 million inhabitants in 2014 (Ji et al., 2014). The development of Shenzhen came with massive foreign investments after it became a special economic zone in 1979. The growth resulted in complex road topology and high traffic demand, leading to traffic congestion problems that propagate over time and space, creating large clusters of congested links (Bellocchi & Geroliminis, 2020; Lopez et al., 2017). The data comprises most of the Futian and the Luohu Districts in Shenzhen, the location of the Central Business District. The considered network consists of 1,858 intersections connected by 2,013 road segments (Fig. 2.1A). As shown in Fig. 2.1A, 50 regions form the demand data. Fig. 2.1C shows the demand in an OD matrix. The colormap represents the frequency of each OD pair in the sample. White points represent OD pairs without entries in the sample.

An MFD represents the traffic congestion and computes the average speeds in the network as a function of the accumulation of private and ride-sourcing vehicles. Note that while speed is represented by an MFD, vehicles are moving following the actual network topology and roads. The MFD used on Shenzhen is based on the one obtained in Ji et al., 2014 for the same data of taxi trips. To approximate the jam accumulation for all moving vehicles in the network, we used the total road length (both ways, in case of multiple lanes) using OpenStreetMap data. We assumed that congestion is homogeneous in the region. Hence, a single MFD is capable of measuring congestion. Another reason for such simplifying assumption is the computation of shortest paths – which remain unaltered during the simulation – and, consequently, the route choice. Eq. [2.1] shows the Accumulation n *vs.* speed v(n) relationship and Fig. 2.1B is the graphical representation. While this simplification created an elegant model with small computational effort, it can still well represent the distribution of trip lengths as in real settings (Section 2.3 provides more details). Other methods for estimating the MFD can be found in Saffari et al., 2022 and Geroliminis and Sun, 2011 and the references therein.

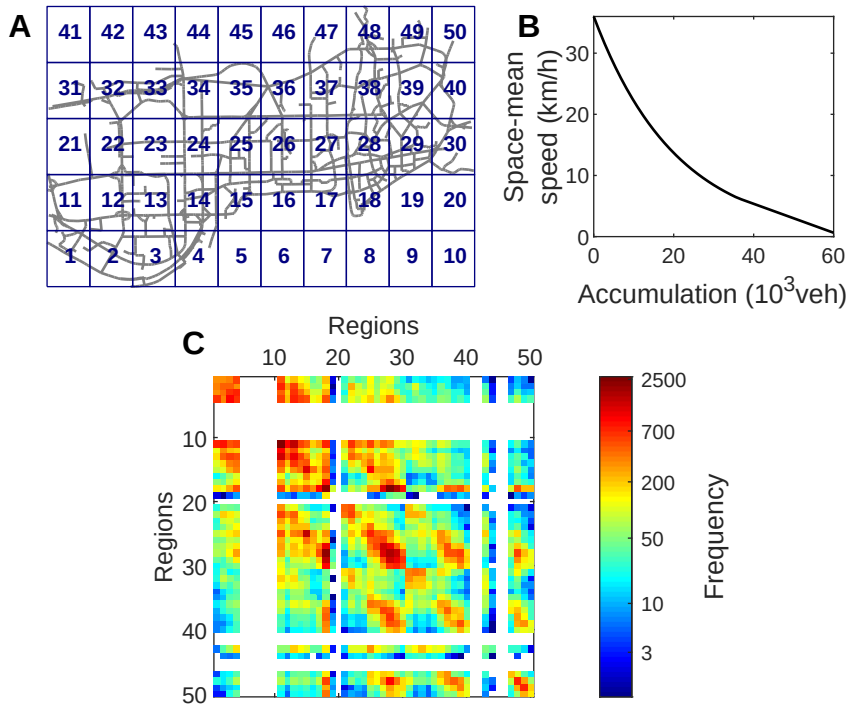


Figure 2.1: The simulation of a ride-sourcing service in Shenzhen used data of detailed network, demand, and traffic to provide accurate measurements on operations of ride-sourcing services. (A) Map of Shenzhen and its demand regions. (B) Space-mean speed vs Accumulation for Shenzhen. (C) Demand density per Origin-Destination pair (in a log scale).

$$v(n) = \begin{cases} 36e^{(\frac{29}{600}m)}, & \text{if } m \leq 36 \\ 6.31 - 0.28(m-36), & \text{if } 36 < m \leq 60 \\ 0, & \text{if } m > 60 \end{cases}, \qquad \text{where } m \equiv \frac{n}{1000} \tag{2.1}$$

2.2.2 State description and Congestion dynamics

Four different entity classes populate the simulation environment: private vehicles (PVs), waiting passengers (WPs), traveling passengers (TPs), and ride-sourcing vehicles (RSVs). Each of the classes has properties to define them, shown in Table 2.1.

Vehicles move in the network following a trip-based model with an accumulation *vs* speed MFD (Arnott, 2013). For every new trip (a PV or an RSV), the model computes its total distance to the destination and updates the remaining distance for each vehicle based on Lamotte et al., 2018. One way to introduce it starts from the simple observation that a vehicle with trip length l_0, which entered at a time t_0, should exit after traveling l_0, i.e., after a time interval τ_0 satisfying Eq.[2.2].

Table 2.1: Nomenclature of tuple elements for each entity.

Entity	Property	Description
PV (Private Vehicle)	PV_i^{id}	Identification
	PV_i^{at}	Arrival time
	PV_i^{o}	Origin
	PV_i^{d}	Destination
	PV_i^{rd}	Remaining distance
WP (Waiting Passenger)	WP_j^{id}	Identification
	WP_j^{at}	Arrival time
	WP_j^{o}	Origin
	WP_j^{d}	Destination
	WP_j^{wts}	Willingness to share
	WP_j^{dr}	Assigned driver ID
TP (Traveling Passenger)	TP_j^{id}	Identification
	TP_j^{pt}	Pick-up time
	TP_j^{o}	Origin
	TP_j^{d}	Destination
	TP_j^{wts}	Willingness to share
	TP_j^{dr}	Assigned driver ID
	TP_j^{td}	Distance traveled
RSV (Ride-Sourcing Vehicle)	RSV_k^{id}	Identification
	RSV_k^{l}	Last passed intersection
	RSV_k^{cd}	Current destination
	RSV_k^{rd}	Remaining distance to the current destination
	RSV_k^{np}	Number of passengers inside the vehicle
	RSV_k^{pID}	ID of assigned passengers (in order of activity)
	RSV_k^{pAC}	List of activities (in order of execution)

$$l_0 = \int_{t_0}^{t_0+\tau_0} v(n(k))\, dk \tag{2.2}$$

The main difference with a classical trip based MFD model with an input the trip length distribution of vehicles (see for example, Lamotte and Geroliminis, 2018), is that we estimate trip length for each trip based on the instantaneous shortest path between origin and destination in the real network. We also estimate the trip length between the points in the network that will change the state of a vehicle as described in more details later in Fig. 2.2. While we are currently using a single MFD for the whole network, this work can be extended in multi-region MFD networks.

The population of PVs fluctuates as every PV has a specific arrival time. Furthermore, we assume that once a PV reaches its destination ($PV_i^{rd} = 0$), it enters a garage or parking lot, leaving the system. Note that RSVs and PVs move in the network at every time step at variable speeds, varying according to the traffic conditions summarized in the MFD (Fig. 2.1B).

WPs are the passenger entities that were not served yet by an RSV. If a WP j is willing to share his ride (hires the ridesplitting service), his willingness to share WP_j^{wts} is set to 1. Otherwise, it is set to $WP_j^{wts} = 0$. The choice for sharing is the result of a single Bernoulli trial for each traveler generated in the system. Nevertheless, it requires a good quality shared service for the system to accomplish it; otherwise, the user will travel alone even in a ridesplitting service. Service quality constraints are described later in Section 2.2.3. Finally, once s/he has an assigned RSV to pick-up, it cannot change, and the property WP_j^{dr} links it to the passenger. Note that WPs may assume two states, waiting for an assignment ($WP_j^{dr} = 0$) and waiting for pick-up ($WP_j^{dr} \neq 0$).

Once an RSV picks-up a WP, the last becomes a TP (leaves the list of WPs and adds a new member to the list of TPs). The new TP inherits most of the data from the WP, said: identification, origin, destinations, willingness to share, and assigned driver. However, TPs have new properties, said: time of pick-up and traveled distance. As the speeds might vary a lot during a trip, a traveling passenger has no information about the delivery time, which is informed once the passenger reaches the destination.

The central entity of the ride-sourcing service is the RSV, which is responsible for the pick-up and delivery of passengers according to their preferences. Different from the PVs, RSVs have their positioning tracked all the simulation long. They also may assume different states depending on their current activity. Every RSV has an identification RSV_k^{id}, a last passed node RSV_k^{l} (node in the network, updated at every time step), a current destination RSV_k^{cd} (node in the network, such as a WP's origin, or a TP's destination, or a parking lot), and a remaining distance to the current destination RSV_k^{rd}, to keep track of their position. To keep track of their activities, they have the destinations' ID RSV_k^{pID} (identification of the passenger – waiting or traveling one), and number of passengers inside the vehicle RSV_k^{np}. We assume that RSVs have a limited capacity of two passengers. A supporting argument for such an assumption is found in W. Li et al., 2019. The authors identified that only 6-7% of trips were shared, and more than 90% of them had at most two passengers in a study in Chengdu, China; shared trips with 3 or more passengers is in the range of 0.7% or even less.

RSVs perform different activities as they move through the network. Fig. 2.2 shows how RSVs change their states during the simulation breaking the activities for both available services. The states refer to the current activity of the RSV. For each activity of an RSV, the path choice follows the Floyd-Warshall algorithm (the shortest path) in terms of distance. In general, an RSV can perform seven different activities:

- Cruising for passenger: the vehicle has no passengers and is driving around his current

location, waiting for an assignment of a new passenger (WP);

- Driving to park or to a hot-spot: the vehicle has no passengers and is driving to a hot-spot near high demand areas and then circulates randomly in this area until a request arrives (in case the parking strategy from Section 2.2.5 is active, the hot-spot becomes a parking lot);
- Parked: the vehicle has no passenger, reached a parking lot near high demand areas, and waits there for the next assignment (only possible when the parking strategy from Section 2.2.5 is active);
- Picking-up a first passenger: the vehicle received the location of a waiting passenger (assignment), and it is moving towards the passenger's pick-up position (origin);
- Delivering a single passenger: after picking-up the passenger, the vehicle drives him/her towards the final destination;
- Picking-up a second passenger (exclusive for ridesplitting): the vehicle has one passenger and is moving towards a second passenger that matched the current ride; and
- Delivering a passenger of a shared ride (exclusive for ridesplitting): the vehicle has two passengers and is moving towards the destination of one of them.

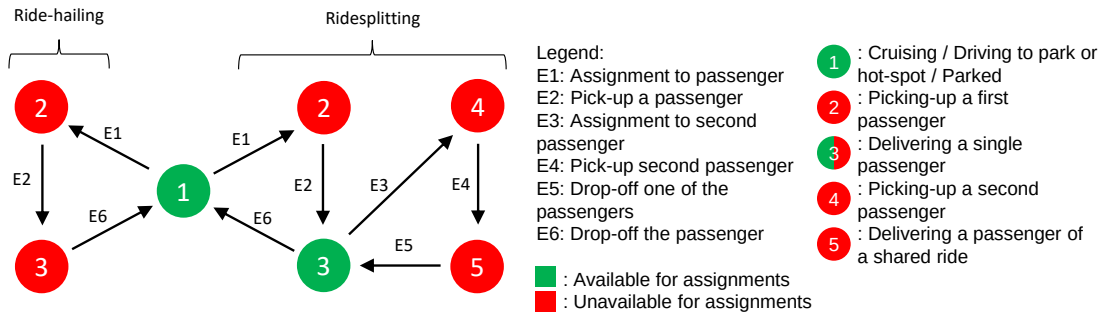


Figure 2.2: Simplified RSV activity flow framework for ride-hailing and ridesplitting services.

2.2.3 Matching passengers and drivers

Here, we present the matching process used inside the simulation and its assumptions regarding driver choice and passengers' matching requirements.

Defining a trip determines the order of the points which the vehicle will visit. The matching process is responsible for determining RSVs' trips. WPs admit waiting 1 minute to receive an assignment (a designated RSV available to pick-him/hr-up) that fulfills the requirements for maximum waiting time and detour. After this time, travelers leave the WP list and choose a new mode of transportation (busses, bike, walk, taxis, private vehicle) in an event called

'abandonment'. Busses, bike, and walking are considered secondary to the accumulation, and, therefore, the simulation does not keep track of their activities. Passengers in these modes transfer with a fixed travel time to their destination. In the case of an abandoning passenger who decides to travel by taxi, it is modeled similarly to PVs. While a mode-choice module could integrate the simulation, this is beyond the scope of the paper that focuses on the supply side. The interest of this work is to analyze the effect of ride-sourcing services in congestion for different fleet sizes and willingness to share. For demand-oriented work, the reader could refer to Tirachini and Gomez-Lobo, 2019, Tirachini and del Río, 2019, Zha et al., 2016, and Wei et al., 2020. The effect of mode choice and socioeconomic characteristics in the ride-sourcing literature is a research priority.

In general, the matching process assigns the RSV with the smallest extra trip length among the five closest RSVs that fulfill all requirements to perform the ride. We define the smallest extra trip length as the distance that the new assignment will make the RSV travel, in addition to any ongoing activity. It means that even an RSV that fulfills all requirements (capable) and may save some more Vehicle Kilometers Traveled (VKT) will not get the passenger if there are other five capable RSVs closer, for instance.

Despite that the system could benefit from optimization in the dispatching and matching processes, we are interested in evaluating such a system as an operation with human agents and their limited rationality (the system is not centrally optimized). For instance, Hanna et al., 2016 points out that services such as Uber and Car2Go assign the nearest vehicle on a first-come-first-served basis. Future research could investigate the effect of more advanced matching optimization techniques on the performance of the system.

Assignment in ride-hailing

Requirements for matching passengers and drivers in ride-hailing derive from our assumptions about passengers' tolerances towards waiting and service definition. The RSV must be idle and able to reach the passenger in less than Δ minutes under current traffic conditions (Eq. [2.3]). Any vehicle that fulfills the latter is capable of serving ride-hailing passengers. The matching process uses the distance between two points ($p(\cdot,\cdot)$) and the current speed ($v(t_{clock})$) to compute its requirements. Recall that RSV_k^l and WP_j^o refer to the RSV's last passed intersection and the passenger's origin, respectively.

$$p(\text{RSV}_k^l, \text{WP}_j^o) \leq v(t_{clock}) \cdot \Delta \tag{2.3}$$

In summary, the matching process for ride-hailing (single rides, without the option to share) searches the closest available RSV. It occurs because the extra traveled distance for idle RSVs is simply the summed distance to pick-up and deliver the passenger.

Assignment in ridesplitting

There is no predefined priority for assigning partially busy RSVs when a ridesplitting passenger arrives. Instead, the matching process looks for the five nearest capable vehicles, including idle ones and those delivering another ridesplitting passenger.

Defining a ridesplitting trip depends on the chosen RSV's current activity. If it is empty, the defined trip is similar to a ride-hailing one. However, when evaluating a shared ride match (a vehicle with one passenger matches with a second passenger), two types of trip schemes arise. Fig. 2.3 illustrates both types of trips (*j-i-j* and *j-j-i* sequences) and direct route (*i-i* sequence). Note that the process only assigns one passenger per run, immediately at the passenger's arrival or when a new RSV becomes available. It means that the algorithm does not wait for a pool of passengers to form after some time. Note that, based on the previous statements, assigning an empty RSV to an arriving ridesplitting passenger will not prevent him/her from sharing the ride later. For this reason, the evaluation of matched rides, such as the one from Fig. 2.3, only happens with an en-route RSV.

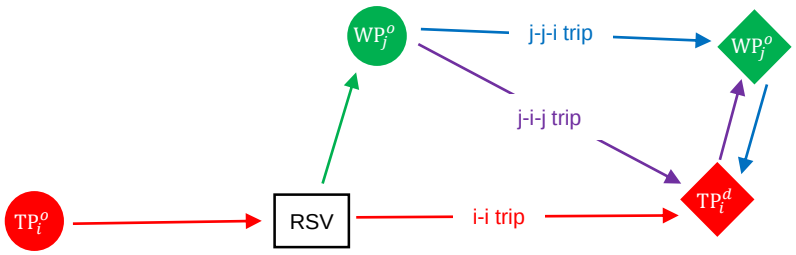


Figure 2.3: Ridesplitting trip options scheme. The 'RSV' box indicates the current position of the vehicle in the illustrative network. '*i-i* trip' refers to a direct trip from TP_i^o to TP_i^d (en-route trip at the moment of the evaluation). '*j-j-i* trip' refers to a ridesplitting trip that will deliver passenger *j* first, and then passenger *i*. '*j-i-j* trip' refers to a ridesplitting trip that will deliver passenger *i*, and then passenger *j*.

A shared trip contains two passengers if both are willing to share[I] and the first one, who is already on-board, has a similar trip (with a small detour) with the second one. Thus, considering all nearest vehicles during the assignment will also potentially increase (when demand for shared rides is high) the number of circulating vehicles with one passenger who is willing to share. This will result in more shared rides during the peak hour when the system needs them. It will also result in better quality of matching because if we force shared rides, this might result in higher waiting times. This is evident in Figs. 2.8 and 2.9, where in the beginning of the simulation the number of ride-splitting with one passenger is more than with two passengers, but as demand increases a higher number of matches occurs and the system performs many shared rides. If matches are forced by only searching the 5 nearest vehicles with already a passenger (who is willing to share), this might create lower quality of service or

[I]Note that we refer to a passenger willing to share as a passenger that already hired ridesplitting service (after a mode choice decision, which is outside the scope of this chapter).

infeasible solutions.

The requirement for an empty vehicle to serve ridesplitting passengers is the same as for ride-hailing (Eq. [2.3]). Additionally, requirements for a shared ride match derive from passengers' tolerances towards deviating from their original path and service definition. Firstly, all involved passengers must have hired ridesplitting rides ($\mathrm{WP}_j^{wts} = \mathrm{TP}_i^{wts} = 1$). In the '*j-i-j*' trip from Fig. 2.3, it is not allowed to add more than a maximum relative detour Ω to the trip distance of TP '*i*'. Thus, the detour of picking-up the WP '*j*' must be acceptable for '*i*' (Eq. [2.4]); the same applies to '*j*' regarding the delivery of '*i*' (Eq. [2.5]). Finally, for the sequence '*j-j-i*', the detour of picking-up and delivering '*j*' must be acceptable for '*i*' (Eq. [2.6]). Note that, in this sequence, there is no detour for '*j*'. In case both sequences ('*j-i-j*' and '*j-j-i*') are possible, the shortest one in distance is chosen.

$$\mathrm{TP}_i^{td} + p(\mathrm{RSV}_k^l, \mathrm{WP}_j^o) + p(\mathrm{WP}_j^o, \mathrm{TP}_i^d) \leq p(\mathrm{TP}_i^o, \mathrm{TP}_i^d) \cdot (1 + \Omega) \tag{2.4}$$

$$p(\mathrm{WP}_j^o, \mathrm{TP}_i^d) + p(\mathrm{TP}_i^d, \mathrm{WP}_j^d) \leq p(\mathrm{WP}_j^o, \mathrm{WP}_j^d) \cdot (1 + \Omega) \tag{2.5}$$

$$\mathrm{TP}_i^{td} + p(\mathrm{RSV}_k^l, \mathrm{WP}_j^o) + p(\mathrm{WP}_j^o, \mathrm{WP}_j^d) + p(\mathrm{WP}_j^d, \mathrm{TP}_i^d) \leq p(\mathrm{TP}_i^o, \mathrm{TP}_i^d) \cdot (1 + \Omega) \tag{2.6}$$

Note that the matching requirements share many similarities with the shareability networks presented in Santi et al., 2014. Moreover, our matching process occurs online, as seen in Alonso-Mora et al., 2017. However, these processes share some key differences: 1) we do not allow to change the vehicle that will pick-up a passenger, and 2) fully occupied vehicles are not options for assignments. Readers can refer to Martinez et al., 2015, Jung et al., 2016, Hosni et al., 2014, Stiglic et al., 2016, Nourinejad and Roorda, 2016, Long et al., 2018, Zeng et al., 2020, Furuhata et al., 2013, and Agatz et al., 2012 for other matching strategies for on-demand transportation services.

2.2.4 Moving idle vehicles to hot-spots

It is unfruitful for RSV drivers to remain in a location that would yield lower revenues. Cruising around the destination of the last assignment might lead to longer vacant times. Hence, we assume drivers prefer to move to areas of high demand after delivering a passenger. Currently, some TNCs test surge pricing schemes to attract vehicles to high demand areas (Lu et al., 2018). Nevertheless, we do not investigate the consequences of surge pricing on ride-sourcing services, drivers repositioning, nor labor supply.

Drivers have *a priori* a list of high-demand areas, called hot-spots, which we define later in this section. We assume that drivers will move to the closest hot-spot once they become idle and then randomly circulate in this region. Once drivers reach the nearest hot-spot, they cruise, awaiting the next assignment. Note that RSVs are available for new assignments during the 'driving to hot-spot' state. Typically, this would not characterize a behaviorally meaningful strategy. However, as seen in Fig. 2.1C, there is a symmetric distribution of origins

and destinations. It implies that drivers have little incentives to move too far from their current position because vehicles naturally achieve a spatial distribution close to the distribution of the origins of new requests. In other words, any additional movement is likely to create a spatial imbalance between available vehicles and incoming demand, such that those drivers that decide to move will likely spend longer periods cruising for the next assignment. We will briefly comment in Section 2.3 the case that drivers do not relocate in hot spots but cruise near the areas of the last drop-off.

The locations of hot-spots are the result of a simplified p-median problem (Owen & Daskin, 1998) in two stages. The first stage defines the intersections with the shortest average distances to other intersections in their respective demand regions (see the 50 demand regions in Fig. 2.1A). The second stage solves the p-median problem for the defined nodes and the demand values for each region. Fig. 2.4 summarizes the positions of hot-spots, their closest intersections, and their demand shares (solution of the p-median problem).

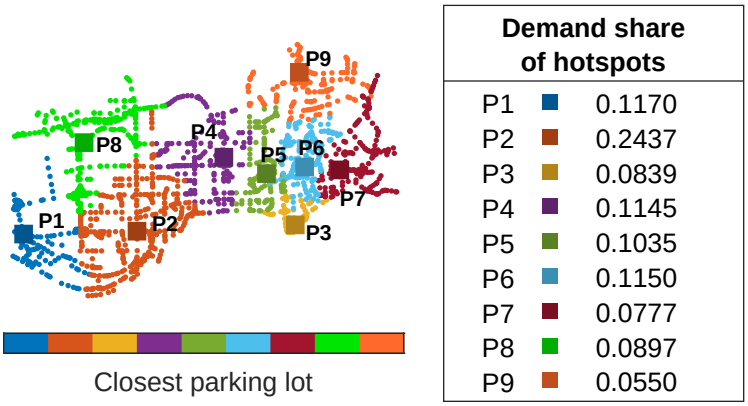


Figure 2.4: Geographical position of hot-spots and the set of nearest intersections. The legend brings the results from the p-median problem (demand share of each hot-spot, used for the parking strategy).

2.2.5 Parking strategy

TNCs' attractiveness depends significantly on the fast response on picking-up passengers when a request arrives (similar to other types of response systems, see, for example, a vast literature for emergency response systems, based on location theory). To succeed in this objective and attract higher demand from other modes of transport, TNCs try to increase the number of registered drivers (see an economic analysis for a static model in Tirachini and Gomez-Lobo, 2019). While an increased fleet size could decrease the waiting time for passenger pick-up, it creates a mass of idle circulating vehicles. As the numerical study of Section 2.3 shows, strong congestion effects might appear in the network. Nevertheless, this congestion affects other modes of transport that move in the same part of the network. Thus, if there is no intervention from the government to penalize the negative externalities of these actions (e.g., through pricing or creating additional opportunities for public transportation),

TNCs can have advantage over other modes of transport. If there is spare parking capacity, the development of simple strategies could decrease the circulation of idle vehicles without significantly increasing the waiting time. The purpose of such is that by leaving on the side demand interactions and mode choice, we can still show with a dynamic model of congestion that smart parking strategies can have a positive effect on the overall system. Whereas the implementation and pricing of such systems can influence mode interactions, this is yet another further research direction in transport economics. This paper focuses more on the supply interactions and dynamics of congestion as a function of fleet sizes and parking strategies that try to decrease the number of circulating idle TNC vehicles.

Idle RSVs cruise, in a random walk, near their last destination waiting for their next assignment. Such behavior has the potential to increase empty kilometers traveled and degrade traffic conditions. For this reason, we propose a parking strategy where we assign idle vehicles to parking lots near high demand areas. The added value of parking lots is that parked vehicles do not contribute to the MFD accumulation and, thus, congestion levels are lower. The idea here is not to provide a fully operational strategy, but to evaluate the potential improvements of preventing empty vehicles from cruising. A detailed evaluation and actual needs of an operational strategy using city-wide parking is a direction for further information, where one must identify how the parking costs are distributed among the authorities, companies and drivers.

Xu et al., 2017 presented an optimal parking provision for ride-sourcing vehicles. The authors focus on managing the trade-off of restraining road capacity, creating curbside parking, and removing idle ride-sourcing vehicles from the streets because it explicitly had cruising RSVs as a source of additional VKT. The framework, however, did not consider ridesplitting nor other spatial network effects directly, such as matching of multiple customers and an RSV.

The locations of parking lots are the same as the hot-spots (from Section 2.2.4). We consider off-street parking lots at these locations. These parking lots have a limited in-park capacity. Therefore, although every idle driver receives a request to move to a parking lot, some of them will not find an in-park spot and will cruise nearby the assigned parking lot. Once the driver reaches the parking lot entrance, the queue to enter follows a first-come-first-served discipline. At the same time, to leave the parking lot, there is a last-come-first-served discipline. Note that RSVs are available for new assignments while in a 'driving to park' or a 'parked' state. The number of in-park spots and assignable spots (in-park spots plus the number of vehicles that will cruise nearby) of each parking lot is proportional to the demand share of the respective hot-spot. We assume, in most scenarios, that in-park capacity is limited to half of the fleet size of RSVs. We discuss in Section 2.3.2 the effect of parking lot capacity in system performance.

Assignment of idle RSVs for parking lots uses a color system to prioritize emptier parking lots. Colors (classifications) are a reference to their usage level, i.e., a parking lot with fewer vehicles have a higher priority to get an assigned vehicle. The parking strategy considers the proximity between RSV and parking lot as a secondary classification. The Drum-Buffer-Rope method

inspired this process (Cox III & Schleier, 2010). The highest priority goes to green parking lots, which have more than 70% of available spots. Then, yellow ones have more than 30%, but less than 70%; red ones have less than 30%; and, finally, black ones (lowest priority) have all of them assigned to drivers. In summary, the process sends drivers to the nearest parking lot with the highest priority at the moment of the decision (flags change according to the instantaneous number of available assignable spots of each parking lot). Section 2.3.2 provides the dynamics of vehicle occupancy in parking lots during the simulation. Note that drivers can receive a request to move to the area of a full parking lot (represented with a 'black' flag). These drivers will cruise near the parking lot until they receive a new assignment (passenger), or a in-park spot becomes available.

2.2.6 Congestion and transit

Passengers' choice for ride-sourcing service depends on waiting times, fares, and journey duration. Once congestion rises, journey duration becomes less appealing. In such situation, passengers become more likely to change to public transportation.

As mentioned earlier, in Section 2.2.2, we consider that transit move at constant speeds as there were dedicated bus lanes that did not interact with traffic. Another approach would have been to utilize a 3D MFD to model car-bus interactions (similar to Geroliminis et al., 2014), but this is beyond the scope of this work. A passenger can change her choice in case travel times become too long in ride-sourcing. The estimates for journey duration in ride-sourcing and transit consider a waiting time and a time inside the vehicle. Every passenger computes these estimates before joining the system. In case transit presents a lower estimate for journey duration, some passengers change transportation mode. Hence, when the system becomes very congested, more passengers might shift to public transport. Differently of abandonments, these passengers do not wait for a driver. They leave the ride-sourcing system immediately at their arrival. One can expect that this takes more passengers from ridesplitting than from ride-hailing since we are only accounting for traveling times, not costs. The opposite is expected from waiting time abandonments, since more vehicles will be available for a ridesplitting service.

Eq. [2.7] computes the estimated journey duration of transit. One can relate it with what a passenger expects to wait plus the amount of time walking plus the transit time-table assumed to travel around a u_{cr} = 10 km/h (walking and in-vehicle average). Although ridesplitting passengers can accept up to the maximum detour tolerance (Ω), they expect only half of it to be used (we show evidence for this assumption later in Fig. 2.7B). Eq. [2.8] computes the estimated journey duration of a ride-sourcing passenger based on current traveling speeds in the network ($v(n(t))$). One can relate such an estimate with what customers see on the application screen before booking a ride. Note that the passenger has a predefined willingness to share, e.g., the service s/he will hire.

$$T_{\text{transit}} = E\left[W_{\text{transit}}\right] + \frac{p(\text{WP}_j^o, \text{WP}_j^d)}{u_{cr}} \tag{2.7}$$

$$T_{\text{TNC}} = E\left[W_{\text{TNC}}\right] + \frac{p(\text{WP}_j^o, \text{WP}_j^d)}{v(n(t))} \times \left(1 + \text{WP}_j^{wts}\frac{\Omega}{2}\right) \tag{2.8}$$

2.3 Computational results

We analyzed several metrics, with different ride-sourcing fleet sizes (from 1000 to 7000 in increments of 500 vehicles – based on the operating number of taxis in Ji et al., 2014 for Shenzhen), and willingness to share (fraction of passengers hiring ridesplitting: 0%, 30%, 60%, and 90%). Among the metrics we analyzed, we can highlight passengers' waiting times, the time a passenger stays in the system, trip lengths for pick-up, delivery and cruising, average traveling speeds, vehicle kilometers traveled, and many others. We assumed that the number of drivers remains unchanged for the time of a single simulation run based on the findings of Zha et al., 2018, where drivers might spend more than six hours working per day (more than double of a simulation run). However, the number of drivers can change daily. Section 2.3.3 presents a discussion in this direction. Further research efforts could analyze within-day dynamics on the number of active drivers, which requires a more careful analysis of pricing and investigation of equilibrium and the trade-off between supply and demand. Note that, in the case of such an application, we expect that fewer drivers would join in off-peak hours. However, the demand peak would surely bring more drivers, specifically if surge pricing schemes are applied, causing similar congestion issues.

The maximum waiting time, Δ, and the maximum detour, Ω, were set to 10 minutes and 20% of the trip length (shortest path from origin to destination), respectively. Expected waiting time for buses ($E\left[W_{\text{transit}}\right]$) is also set to 10 minutes according to Fu et al., 2020, while the expected waiting for ride-sourcing ($E\left[W_{\text{TNC}}\right]$) is pre-computed using the simulation. From all waiting abandonments, about half choose to travel by busses or bike or walk. The other half call a taxi or pick-up a private vehicle (see Rayle et al., 2016). As the number of abandonment trips is small for fleet sizes above 2000 vehicles (1-4% from all trips, including PVs and ride-sourcing), variations in the fraction of these trips between public and private modes do not influence the numerical results, and conclusions remain unchanged. There are separate scenarios to evaluate the parking strategy from those where it is deactivated.

A Poisson process describes the arrival process of both PVs and WPs. They had piece-wise constant rates in a 3-hour long simulation with a low-high-low demand profile lasting one hour for each period. In the low-demand profile, private vehicles and ride-sourcing passengers split in 34'000 and 6'000 trips per hour for each, respectively. In the high-demand profile, arrival rates double.

Our results show that encouraging ridesplitting is not enough to decrease the VKT, a measure

associated with worse congestion, fuel consumption, and safety issues. Furthermore, traffic congestion worsens as ride-sourcing fleets grow. Finally, the findings acknowledge that it is necessary to restrain idle ride-sourcing vehicles from cruising to decrease impacts on VKT.

As this is a trip based simulation, with an MFD representation of speed dynamics in the network, it is necessary to test whether this parsimonious model, without link speed variations and detailed traffic assignment, provides realistic traffic characteristics. To do so, we compare trip length distributions from the real taxi trips in Shenzhen with the ones produced by the simulator. We sampled 2'000 trip lengths between 7:00 and 10:00 from the taxi trips with passengers from Shenzhen and 2'000 trips from an instance of the simulation, to ensure that the simulation could provide a realistic representation. Fig. 2.5 summarizes the probability density functions for both samples and compares their cumulative density functions. Graphically, both samples have similar shapes. Furthermore, a Kolmogorov-Smirnov test evaluated the similarity between the samples and did not reject the null hypothesis for a confidence level of 5%. The use of the shortest path and aggregating demand data in regions (see Fig. 2.1A) did not generate a significant distinction between samples such that the simulator could represent trip lengths accurately.

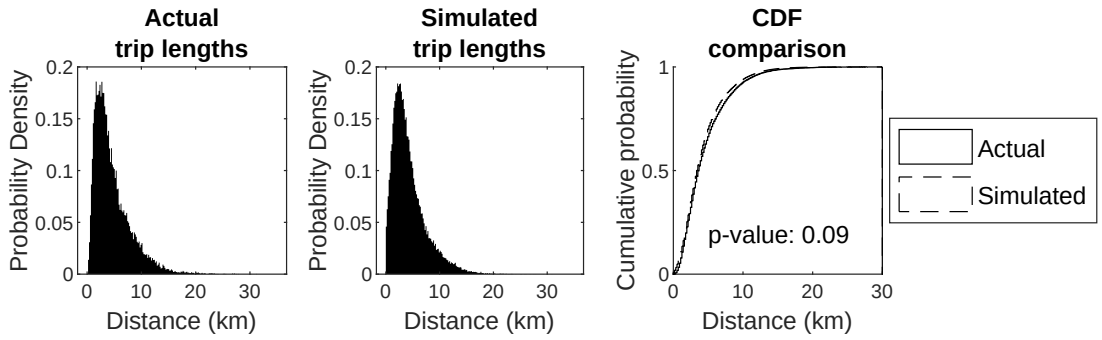


Figure 2.5: Histograms and CDF of trip lengths. Result of a Kolmogorov-Smirnov test comparing both samples.

2.3.1 The effect of willingness to share and fleet size on service quality

Evaluating the ride-sourcing service requires a multi-dimensional look. Performance measurements include waiting times, journey duration, and abandonments regarding the perspective of passengers. While the fraction of travelers that are willing to share a trip is an input to the simulator (ranging from 0 to 90%), the quality of the matching could influence the actual number of travelers that share a trip. It depends on the extra detour for both travelers that potentially match, as described in Section 2.2.

We consider that a complete journey of a passenger starts the moment s/he orders the service and ends the moment s/he reaches his destination. However, passengers that abandon the ride-sourcing service may lead to unrealistic results. Furthermore, waiting and traveling times may underestimate the consequences of abandonments since served trips will concentrate

near main demand centers, whereas those far from them will abandon unserved. For instance, abandonments range between 15% and 33% (for willingness to share of 90% and 0%, respectively) for a fleet size of 1500 ride-sourcing vehicles and decrease to negligible values for larger fleets. For this reason, abandonments penalize the measurements proportionally to the ratio of abandoned passengers using Eqs. [2.9] and [2.10].

$$E\left[W'\right] = E\left[W\right] \cdot (1 + f_{ab}) \cdot (1 - f_{trs}) + \Delta \cdot f_{trs} \tag{2.9}$$

$$E\left[T^{jr}\right] = \left(E\left[W\right] + E\left[T^{tr}\right]\right) \cdot (1 + f_{ab}) \cdot (1 - f_{trs}) + E\left[T_{\text{transit}}\right] \cdot f_{trs} \tag{2.10}$$

Here, $E\left[W'\right]$ represents the average waiting time subject to a penalty, and $E\left[T^{jr}\right]$ describes average journey duration calculated from the average waiting time ($E[W]$), the average travel time ($E\left[T^{tr}\right]$), the waiting tolerance (Δ), the average transit journey duration ($E\left[T_{\text{transit}}\right]$), the abandonments (as a ratio f_{ab} ranging between 0 and 1), the losses to transit (as a ratio f_{trs} ranging between 0 and 1).

As mentioned, the quality of service of transportation services (TNCs, taxis, metros or buses) significantly depends on the passengers' waiting times, which is zero for private cars. Bringing passengers into the service requires planning on the business model, comprising fleet sizes, service availability, fares, so on. An on-demand transportation service, such as a ride-sourcing service, has to manage the dispatching process of its fleet in real-time, accounting for the route choice and chances to match passengers. Operators may reposition drivers establishing fares dynamically through the city. Once dispatching and repositioning policies are well defined and operational, decreasing the waiting times of passengers requires increases in the fleet sizes unavoidably. For example, a passenger may wait between 4 and 9 minutes when only 1000 vehicles are operating (Fig. 2.6A). For some fleet sizes, a passenger may wait between 2 and 3 minutes, on average. Such short waiting times can make ride-sourcing services more appealing compared to public transport (Hensher & Rose, 2007). This can be problematic from a system's point of view because the minimum waiting times occur in fleets larger than those that minimize the average journey duration, and the difference is higher for lower willingness to share. Such an inconsistency can become a problem because companies might prefer to have larger fleets than the optimum (in terms of average journey duration), so they can attract more customers with lower waiting times and increase their revenues. We must point here that it would only materialize depending on the TNCs' capacity of attracting drivers, which need an admissible income from this activity. However, the penalties on the average waiting times increase more due to raises in 'losses to transit' (f_{trs}), than it decreases due to the additional drivers becoming closer to passengers, at large fleet sizes. Note that these losses are a consequence of ride-sourcing becoming less attractive because passengers expect low traveling speeds (see Eqs. [2.7] and [2.8]). It is interesting to observe that nearly the same fleet sizes minimize both waiting times and journey durations when the parking strategy is active.

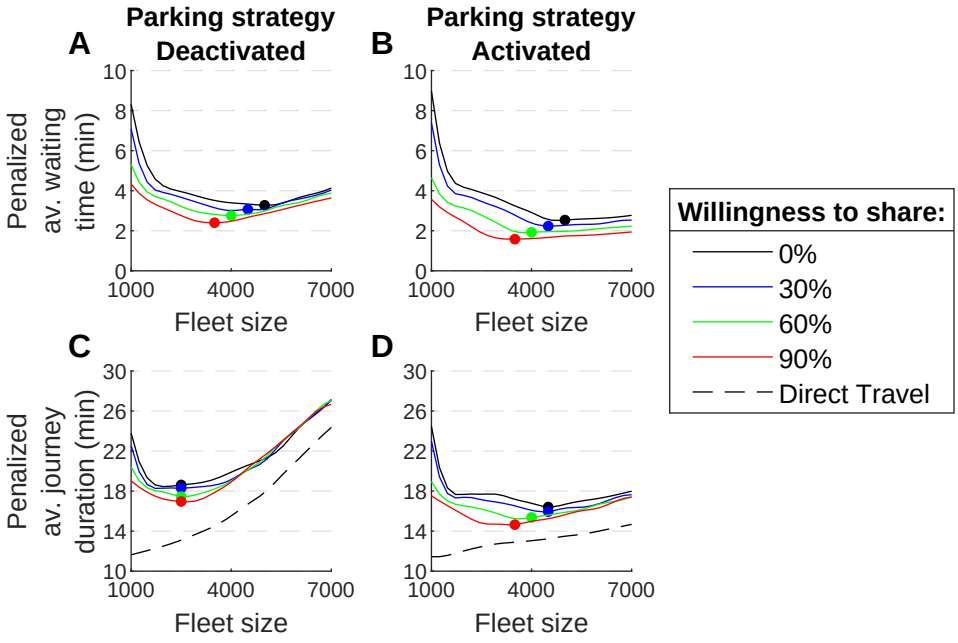


Figure 2.6: Optimum fleet sizes minimize the average journey duration of passengers. (A, B) Series of average waiting times for different fleet sizes and willingnesses to share. (C, D) Series of average journey duration. A and C show results for instances without use of the parking strategy. B and D show results for instances using the parking strategy. Markers indicate the fleet size with minimum journey duration. The results were corrected with a penalty to abandonments. Direct travel time represents the in-vehicle time for a direct service with no detour.

It is interesting to indicate that scenarios without vehicles moving to hot-spots (not shown in the figures) had lower waiting times than the scenarios shown in Fig. 2.6A. They ranged lowered up to values under one minute (lower than the minimum when using the parking strategy). The reason for such finding is in Fig. 2.1C. Origins and destinations have high correlation, thus, remaining near the last delivery point created decent chances of getting a new assignment.

After picking up a passenger, the ride-sourcing service must plan and execute the delivery. Planning includes setting a proper path/route according to a specific strategy, such as minimize travel time for the driver. As a trip-based single region MFD dynamic model is used, once establishing the shortest path, traveling times would only decrease upon higher movement speeds, but the routes would remain unchanged. There resides the conflict of managing fleet sizes for ride-sourcing services. Since traveling speeds depend on traffic conditions, increasing fleet sizes significantly influence average travel speeds. For instance, Fig. 2.6C presents the result of combining shorter waiting times and longer traveling times, producing a well-defined minimum on each case. We define with a marker the fleet sizes at the minimum average journey duration as optimum fleets. Note that higher willingnesses to share lowered the average journey duration and the fleet sizes in the optima. The line “direct travel” indicates

the in-vehicle travel time for immediate service, excluding waiting, detour, and abandonment penalty. It represents the aggregated congestion model of the city, based on the MFD of Fig. 2.1D, and it emphasizes the importance of integrating a congestion model in the analysis. Movie S1 shows how the identification of the optimum fleets relates to the number of idle vehicles at peak-hour getting close to zero.

The parking strategy helps to control congestion and to avoid unnecessary vehicle presence in the streets. The average journey duration and waiting times reach a minimum, and they rise slowly for larger fleet sizes (Fig. 2.6B and 2.6D). For instance, observing the minimums marked with a circle in Fig 2.6, the activation of the parking strategy lowered the average journey duration by 2.3 minutes, for a willingness to share of 90% (red circles in Figs. 2.6C and 2.6D at fleet sizes of 2500 and 3500 vehicles, respectively). With fewer vehicles on the streets (but available in a parking lot), distances to pick-up passengers did not increase, and average waiting times became approximately 1 minute shorter for all values of willingness to share (comparing the minimums – circles – in Fig. 2.6A and 2.6B).

Ridesplitting has two major uncertainties when trying to get more customers. On the one side is the uncertainty of matching passengers, while on the other side is the extra time and distance that they will deviate from their initially designed trip. In Fig. 2.7 we explore these concerns in a growing fleet size perspective. Although small fleets may provide higher chances for matching passengers, they face long detours and waiting times that increase abandonments. The reasoning behind such result is simple. For a given demand, having smaller fleet size means that serving vehicles will be very busy. As a result, the number of vehicles with no passengers will be smaller and in some cases when a new demand arrives, only vehicles with passengers might be in the proximity, pushing the system to create more matches to avoid abandonments. In summary, when fleet sizes are small, customers of ridesplitting have higher chances of being served because of the shortage of empty RSVs. On the other hand, large fleets provide the opposite situation without significant decreases in the matched rides, showing that the matching process indeed favors shared rides to save some VKT. Additionally, the detours decrease for larger fleet sizes because the matching algorithm could match passengers with more similar requests more frequently. All the previous indicates that fleet size is vital in identifying potential matches. Note also that because abandonment rate decreases with fleet size, there is a slight increase for large values due to high level of congestion. The loss of RSV demand to public transport happens only for very large fleet sizes when the network reaches high level of congestion.

To understand these outcomes, we need a more extensive examination of these instances. With a fixed fleet size, we can observe the situation of ride-sourcing vehicles through time and evaluate how it may influence service performance. Fig. 2.8 illustrates the system conditions through the number of vehicles in each state. One can readily note the effect of the peak-hour over the system. Firstly, idle vehicles rapidly become busy. Secondly, in their absence, the number of vehicles working with shared trips ('2nd pick-up' and 'Drop-off shared' states) rises substantially. The parking strategy holds empty vehicles from cruising, accelerating the system

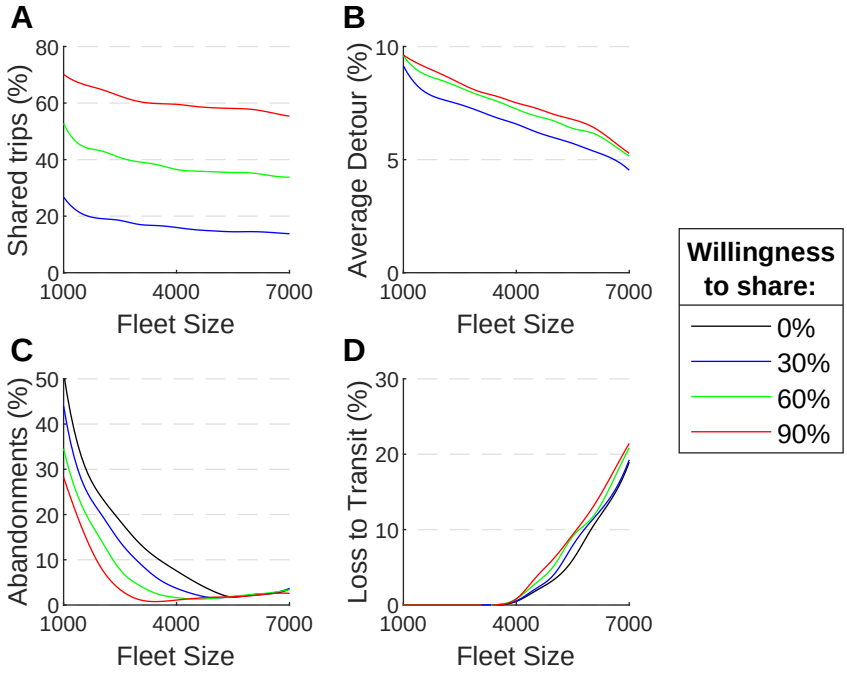


Figure 2.7: (A) Shared trips fraction (accounting only for ridesplitting hired trips), (B) average detours for increasing fleets, (C) fraction of abandonments, and (D) loss to public transport.

recovery from the peak hour. Note that, during off-peak hours, a few vehicles were allowed to cruise without compromising service performance (from Fig. 2.6). The use of the parking strategy enabled shared rides earlier than in scenarios where is is deactivated (see 'Drop-off shared' state when willingness to share is 90%). However, it did not succeed in fostering more shared rides. Table 2.2 shows the VHT used to pick-up and Drop-off shared rides is smaller, especially for larger willingness to share. In general, VHTs for busy vehicles do not change in order, but they fall about four times when observing empty vehicles ('Cruising' and 'Driving to park' states).

System congestion may produce changes in the participation of each service (ride-hailing and ridesplitting) in the number of served passengers. Fig. 2.9 illustrates the allocation of passengers between each service, highlighting shared rides of ridesplitting for scenarios with a fleet size of 3000 vehicles. Note that we did not plot the results for willingness to share of 0% because all passengers hire ride-hailing services (without ridesplitting option). Mainly, ride-hailing demand remains constant for most of the simulation time. The exception is the peak-hour, where ride-hailing loses space for ridesplitting beyond the willingness to share. The reduction in ride-hailing trips occurs between 1.5h and 2h when demand is high, and the number of idle vehicles is small (see Fig. 2.8). At the same time, only shared rides become available, another reason for increasing their proportions. Noticeably, shared ridesplitting rides build up once the matching algorithm is capable of finding partially busy vehicles. Thus, shared trip percentage does not increase instantly with the increase in demand. The later

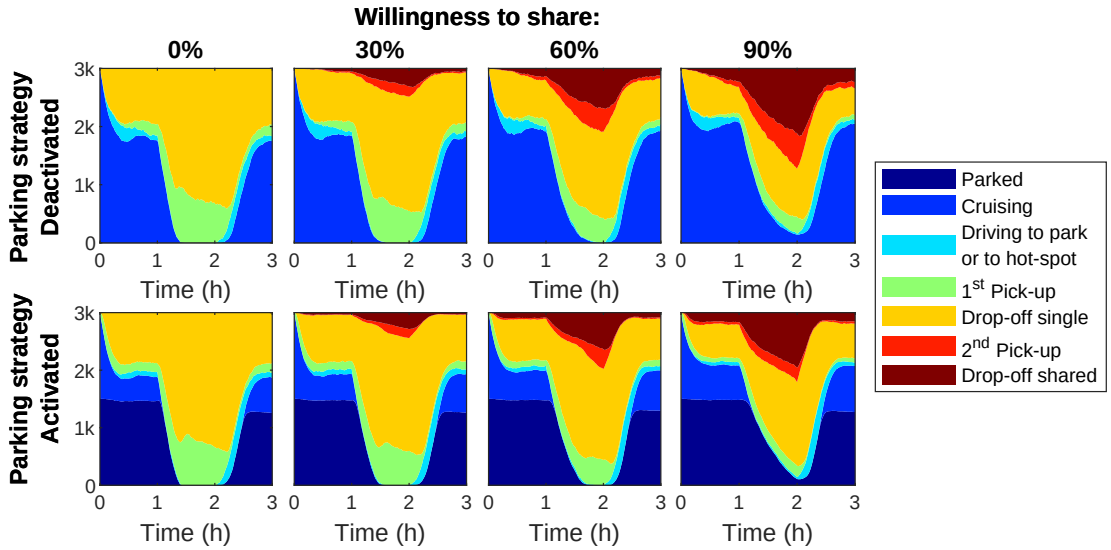


Figure 2.8: Number of vehicles in each state for instances with a fleet size of 3000 ride-sourcing vehicles, and varying willingness to share and use of the parking strategy.

indicates that shared trips require a pool of vehicles with a single passenger to form before starting to share. Once the pool of drivers for ridesplitting is big enough, about half of arriving ridesplitting customers have a shared ride. This illustrates the potential VKT savings at a ride level, from the perspective of the matching algorithm.

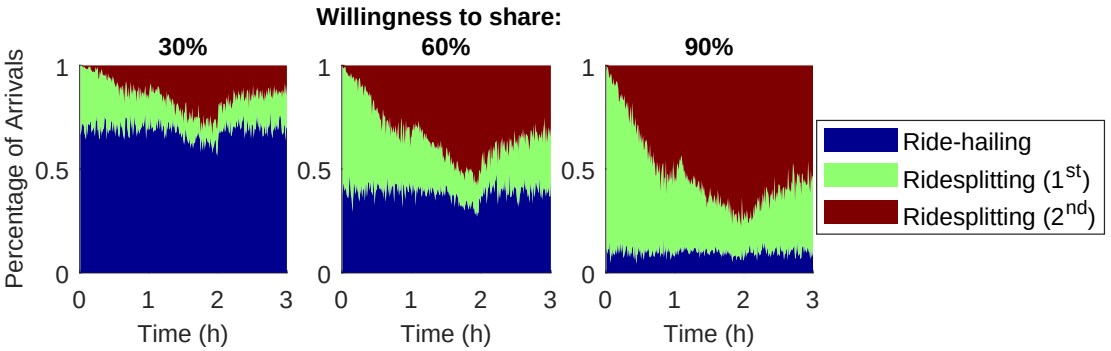


Figure 2.9: Instantaneous percentage of arrivals (out of all served requests) for each service (ride-hailing and ridesplitting) at different willingness to share. Scenarios with a fleet of 3000 ride-sourcing vehicles.

2.3.2 Traffic and TNCs relation

Ride-sourcing vehicles compose urban traffic, influencing traffic performance depending on their actions. Fig. 2.10 reveals that traveling speeds take longer times to recover from the peak-hour for larger fleet sizes. Parking idle ride-sourcing vehicles enhanced the recovery speed from the peak-hour, therefore, increasing the overall resilience of the system (L. Zhang

Table 2.2: Vehicle Hours Traveled (VHT) for private vehicles and for each state of ride-sourcing vehicles (fleet size of 3000 vehicles).

Parking strategy	Willingness to share	Private vehicles	VHT per state					
			Cruising	Driving hot-spot	1st pick-up	Drop-off single	2nd pick-up	Drop-off shared
Deactivated	0%	31503	3144	350	1066	4439	0	0
	30%	31192	3282	362	889	3928	189	350
	60%	30528	3597	377	638	3136	401	852
	90%	29957	4054	376	382	2253	566	1370
Activated	0%	29475	802	297	1069	4270	0	0
	30%	28466	862	304	912	3879	126	290
	60%	28046	962	300	649	3362	227	701
	90%	26995	1156	299	371	2696	264	1064

et al., 2019). Under a fleet size of 3000 ride-sourcing vehicles, instances without the parking strategy reached the critical speed (speed which maximizes flow in the MFD) 35 minutes after the start of the peak-hour and entered a hyper-congested state for another 35 minutes. On the other hand, with the same fleet size with the parking strategy active, they reached the critical speed after 41 minutes after the beginning of the peak-hour and entered a hyper-congested state for 26 minutes. Furthermore, the speeds worsen faster for larger fleets when the parking strategy is inactive.

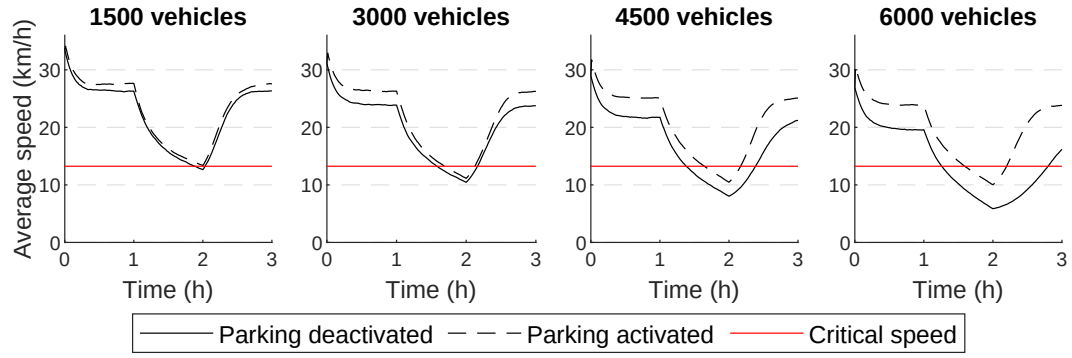


Figure 2.10: Growing fleets deteriorate average speeds and their restoration after the peak-hour. Parking idle vehicles have enhances average speeds independently of the fleet size.

Next, we explore, in Fig. 2.11, the reachable area from an intersection in the central business district as a function of time. As time advances, the driver can travel longer distances until reaching the whole network modeled. A 0.5-hour difference in departure time changes significantly in the reachable area. For example, a driver departing 1.5 hours after the simulation start can reach a distance of 5.4 kilometers, comprising 48% of the simulated network, in 30 minutes, and this reachable area will extend to 60% if departing 0.5 hours later. However, traffic conditions recover faster when using the parking strategy, allowing a driver to travel

4.6 kilometers more by departing 0.5 hours later, whereas only 3.0 kilometers more without the parking strategy (45 minutes travel). Movie S2 shows the evolution of reachable areas and speeds in the simulation.

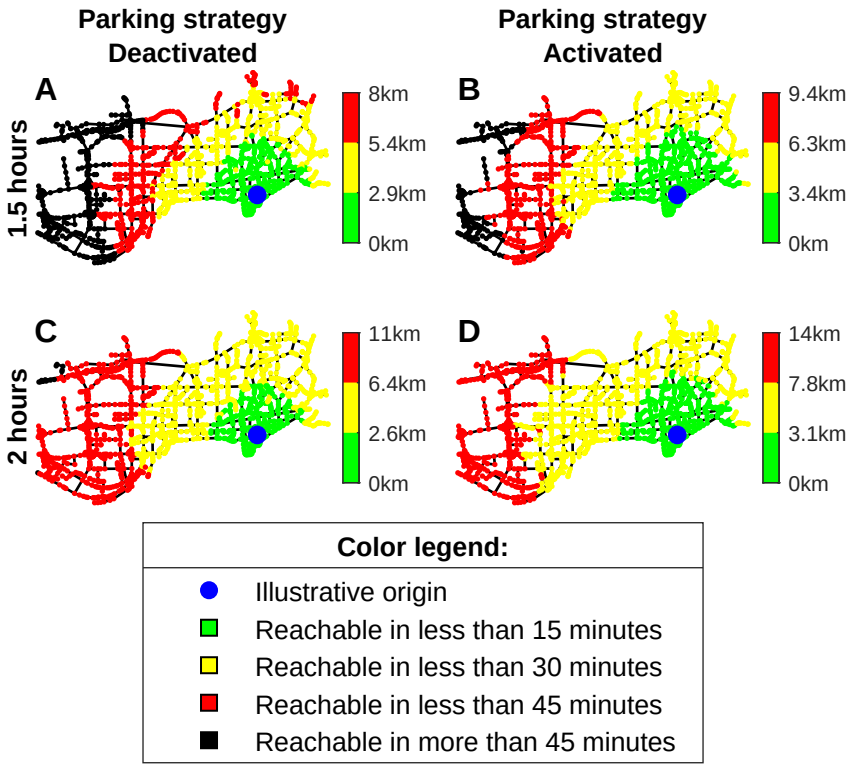


Figure 2.11: Taking idle vehicles from the streets enhances the area a traveler can reach in a certain time window and enhances the resilience of the system, in general. Dynamic reachable area in the modeled network (scenarios with 4000 ride-sourcing vehicles and 30% willingness to share). Starting from an intersection in the central business district of Shenzhen (marked as a blue circle), the reachable area (and distance) that one can access within a certain time window (i.e., 15 min, 30 min, and 45 min) at certain simulated times (A and B 1.5 hours after simulation starts, (C and D) 2 hours after simulation starts). See Movie S2 for a complete observation of reachable areas over time.

Fig. 2.12 shows the number of assigned vehicles to each parking lot. Firstly, most parking lots empty following a similar trend until the point they reach a green flag. At peak-hour, all except parking lot P9 reach this flag. Parking lots P3 and P5 have no vehicles around 2 hours after the simulation start. Except for parking lot P9, which kept nearly 30% of its capacity, very few vehicles remained parked. Yet, this strategy could yield positive results for traffic conditions (Figs. 2.10 and 2.11). The period after peak-hour illustrates the effects of the parking assignment algorithm, where those parking lots that reached the flag thresholds have a lag before receiving new assignments. Only parking lot P5 did not have drivers waiting for a parking spot (black flag) after the peak hour. The opposite happens to parking lots P8 and P9. They are far from the central business district, and, thus, most of the drivers around them

were idle by the end of the simulation.

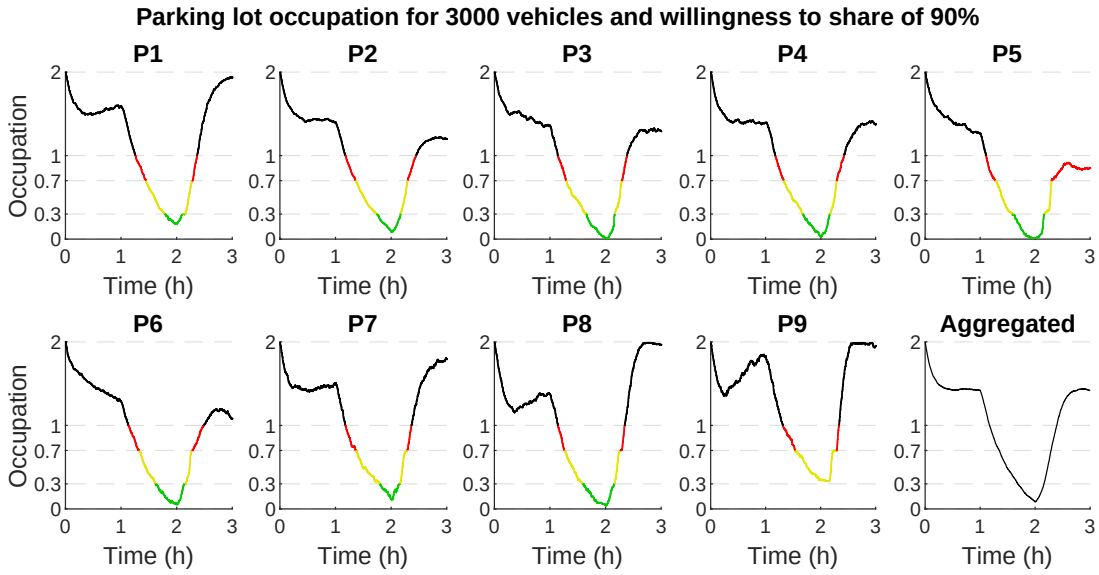


Figure 2.12: Parking lot occupation level and their instantaneous color flags.

VKT is a fundamental measure for transportation systems since it is associated with worse congestion, fuel consumption, and safety issues. Ride-sourcing vehicles generate VKT not only when transporting passengers, but also when they are searching for them, resembling taxis. Hence, we explore through Fig. 2.13 how ride-sourcing fleets generate additional VKT to the city. In a scenario with enlarging fleets of ride-sourcing, the only alternative to curb the growth of VKT is to take idle ride-sourcing vehicles from the streets, as, for example, with the parking management strategy. We show on Eq. [2.11] how ride-sourcing generated additional VKT (VKT$^+$) compared to a scenario where all travelers would travel alone to their destinations without the service (assuming no issues with parking).

$$\mathrm{VKT}^{+} = \underbrace{\int_{t_i}^{t_f} \left(N_{\mathrm{RSV}}^{NP}(t) + N_{\mathrm{PV}}(t)\right) v(t)\, dt}_{\text{VKT of all vehicles}} - \underbrace{\sum_{\mathrm{TP}} p\left(\mathrm{TP}_j^o, \mathrm{TP}_j^d\right)}_{\substack{\text{Sum of passengers'} \\ \text{shortest paths}}} - \underbrace{\sum_{\mathrm{PV}} p\left(\mathrm{PV}_i^o, \mathrm{PV}_i^d\right)}_{\substack{\text{Sum of private vehicles'} \\ \text{shortest paths}}} \tag{2.11}$$

Here t_i and t_f indicate the beginning and the end of a simulated instance. $N_{\mathrm{RSV}}^{NP}(t)$ refers to the number of ride-sourcing vehicles on the streets (not parked), N_{PV} refers to the number of private vehicles on the streets, and $v(t)$ refers to the speed on the network at time t. $p(\cdot,\cdot)$ is the shortest path distance between two points in the network. TP is the group of all served passengers of the ride-sourcing service, and TP_j^o and TP_j^d are the origins and destinations of

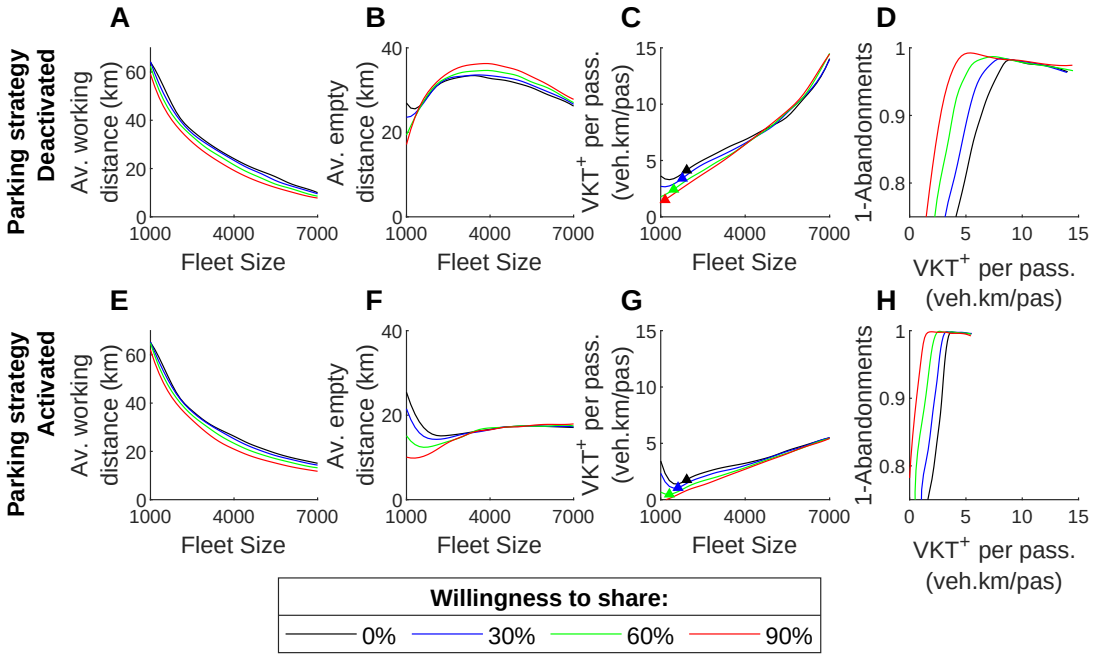


Figure 2.13: TNCs can increase the chances of serving passengers at the expense of adding extra kilometers traveled to the system. (A and E) average working distance traveled per RSV. (B and F) average empty distance traveled per vehicle (C and G) Relationship between the additional kilometers traveled per served passenger and the fleet size. (D and H) Fraction of served passengers as result of additional VKT. All VKT measures are normalized by the fleet size or the number of served passengers. Triangular markers in C and G indicate the fleets capable of serving 75% of the passengers for each willingness to share.

a passenger j, respectively. PV is the group of all travelers that used private cars (or taxis in case of abandonments), and PV_i^o and PV_i^d are the origins and destinations of a traveler i from this group, respectively. Remember that, private vehicles are assumed to use the shortest path, and to leave the system once reaching their destinations, thus Eq. [2.12] holds.

$$\int_{t_i}^{t_f} (N_{\mathrm{PV}}(t))\, v(t)\, dt = \sum_{\mathrm{PV}} p(\mathrm{PV}_i^o, \mathrm{PV}_i^d) \tag{2.12}$$

Thus, Eq. [2.11] may be simplified to Eq. [2.13].

$$\mathrm{VKT}^+ = \int_{t_i}^{t_f} N_{\mathrm{RSV}}^{NP}(t)\, v(t)\, dt - \sum_{\mathrm{TP}} p(\mathrm{TP}_j^o, \mathrm{TP}_j^d) \tag{2.13}$$

We separated the VKT that RSVs produced according to their activities. For instance, 'Working

VKT' refers to those produced when RSVs are assigned to pick up passenger(s) or driving with passenger(s) to destination(s); 'Empty VKT' refers to those produced when RSVs are circulating without a passenger on-board or without a request to pick-up a passenger. Then, we plot on Fig. 2.13 the relation between the fleet size and the Working VKT and Empty VKT divided by the fleet size. The latter variables are equivalent to the average distance traveled per vehicle during the simulation being in one of 'working' or 'empty' states (recall Fig. 2.2). We also plot the additional VKT (VKT^+) per passenger served as a function of fleet size and the abandonment rate. More specifically, on Figs. 2.13A and 2.13E, one can note that RSVs travel approximately the same working distance for a given fleet size with or without the parking strategy. Moreover, in both cases, a smaller value of Working VKT per vehicle (meaning average working distance) is observed as the fleets become larger, which can have an impact on the revenues (as we discuss in Section 2.3.3). On the other hand, the parking strategy had a significant impact on the amount of Empty VKT that each vehicle produced, i.e., average distance without passengers traveled per vehicle during the simulation (Figs. 2.13B and 2.13F). RSVs travel longer distances without passengers while fleets grow until congestion decreases traveling speeds. It is worth noting that, for small fleets (1000 vehicles, for instance), there is the opposite effect, which is related to Wild Goose Chase effect, even with a limited search radius (as suggested in Xu, Yin, and Ye, 2020). Moreover, activating the parking strategy keeps the empty distance traveled near 18km for all willingness to share. Figs. 2.13C and 2.13G show the relationship between growing fleet sizes and the added VKT to the system per passenger served. At the same time, the parking strategy significantly decreased this growth. Note that the markers indicate the fleet sizes capable of serving 75% of the demand and that willingness to share allows fewer vehicles to handle the same number of passengers while producing fewer added VKT. Finally, Figs. 2.13D and 2.13H present an interesting finding related to demand served, which is directly connected to potential benefit for these services. Instead of showing the fleet size, we investigate the relation between served rate (1 – abandoment rate) and added VKT per passenger served. These figures show that there is always a 'price to pay' to serve more passengers, and this is related to further extra kilometers traveled by an increased fleet size. Nevertheless, this graph highlights that significantly higher fleet sizes do not contribute to serving more passengers but only create wasted extra kilometers in the system that result in more emissions, environmental impacts, and safety issues.

To further decrease VKT and VHT, parking lots can increase their capacities. Fig. 2.14 summarizes the effects of parking capacity over the system for fixed fleet size and willingness to share. Note that traffic congestion did not increase sufficiently to make ride-sourcing less attractive than public transport in any of these scenarios. Increasing the number of available parking spots exhibits an option to improve service quality, decreasing waiting times, travel times, and serving more passengers. These improvements result from higher traveling speeds after removing cruising vehicles from the roads (recall the speed *vs* accumulation relationship from the MFD). Furthermore, it improves traffic congestion, reducing VKT. Finally, note that gains in performance appear in even for small capacities. Nevertheless, parking capacity comes with an infrastructure cost that has to be investigated in a future direction (possibly together

with parking pricing schemes).

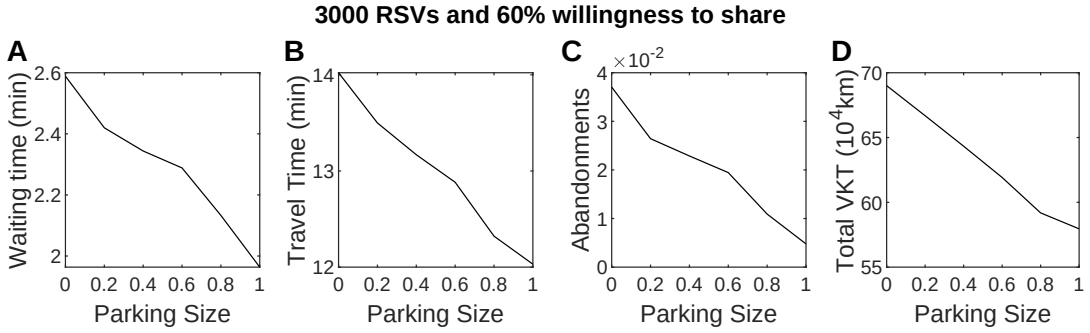


Figure 2.14: Increasing parking capacity (as a fraction of RSVs that can use in-park spot) decreases waiting times (A), travel times (B), abandonments (C) and VKT(D).

2.3.3 Revenues and Day-to-Day Adjustments

In the previous section we considered that the number of drivers is fixed externally and does not change from day to day. Nevertheless, in the supply side of ride-sourcing, there is an inherent feedback mechanism that each driver based on his/her reserve costs and gains from the market join or leave the market dynamically. In other words, the fleet size can vary from day to day. TNC can control the maximum registration of RSV, or apply a cap on maximum active RSV, but the actual active RSV depends on the market, fare, and wage. This section tries to shed some light in this direction by considering a straightforward day to day evolution.

The previous section showed that while larger fleet sizes decrease waiting time for passengers, this creates a higher congestion level and lower quality of service for all private modes of transport. It is known that in competitive markets where different jurisdictions have not the same objective function, the system can reach states far from optimal welfare (see, for example, Douglas, 1972 and Lamotte et al., 2017). While our work does not analyze equilibrium conditions between competitive players, we intend to show that TNCs may end with a pool of active drivers larger than the optimum fleet. However, the negative outcomes of large fleets over congestion will make only a few drivers get a revenue above a minimum wage. Therefore, large fleets do not pose as a potential equilibrium for reasonable minimum wages.

In general, a ride-sourcing service can attract drivers as far as these drivers can profit from offering rides. Both drivers and TNCs can only profit if drivers complete trips. So far, we observed how several settings were able to improve the service perception of customers and their effects on congestion. We assume that the maximum revenue the system can make is to serve all presented ride requests before facing abandonments or loss of attractiveness to public transport. We consider booking fees of US$ 2.20 (F^{rh}_{book}) and US$2.00 ($F^{rs}_{book}$) and fares per kilometer of US$ 1.00 (F^{rh}_{dist}) and US$ 0.80 (F^{rs}_{dist}), for ride-hailing and ridesplitting, respectively (Uber, 2019). Under such assumptions, it is straightforward to compute the potential revenue for the system (R^{p}_{sys}) with the number of passengers (approximately 24'000

passengers before abandonments or losses to transit) and their willingness to share. Equation 2.14 illustrate the estimation of potential revenues.

$$R^{p}_{sys} = N_{req} \cdot \left[(1-\beta) \left(F^{rh}_{book} + E\left[T_{length}\right] F^{rh}_{dist} \right) + \beta \left(F^{rs}_{book} + E\left[T_{length}\right] F^{rs}_{dist} \right) \right] \tag{2.14}$$

Here, N_{req} is the number of requests without any abandonments or losses to transit, β is the willingness to share (in percentage), and $E\left[T_{length}\right]$ is the average trip length of passengers at their shortest path.

The service faces inefficiencies and randomness that reflect in the difficulty of producing the potential revenue. Although the inefficiencies (such as service design, internal policies, and other subjective activities) are hard to enumerate completely, one can do it for most of its impacts over the final revenue. In Fig. 2.15, we explore the revenues the system produces in a single day, as a function of fleet size. Revenues peak at smaller fleet sizes once willingness to share increases (Fig. 2.15A). However, a lower willingness to share yields higher revenues. With an inactive parking strategy, the revenues start to drop immediately after the peak. Note that the peaks occur between 3500 (90% willingness to share) and 4500 (0% and 30% willingness to share) vehicles, depending on the willingness to share. Fig. 2.15C clarifies the impacts of inefficiencies over the potential revenues showing the losses and actual system revenues. For smaller fleet sizes, most of the losses are due to the incapacity to serve all passengers, that finally abandon the system. However, as more drivers become available, they cover these losses up to the point which the system starts to lose demand for transit. Hence, one can expect that the ridership of transit would increase as the number of drivers for ride-sourcing services rise. The 'Incomplete trips' entry stands for the revenue from ongoing rides at the end of the simulation. It is arguably a loss since the system will produce the revenue once these passengers are delivered. However, its growth for large fleet sizes exposes the problems with longer journey durations. Revenues for the system peaks and remains when the parking strategy is active because it avoided the losses to transit, even when fleet sizes were large. However, one can expect revenues losses to transit in case parking lots were smaller. It should be highlighted that we do not propose to increase the fleet size of RSVs to attract more passengers to public transport, as this creates a problematic state for the system with high congestion.

Hall et al., 2018 showed the ambiguous effect of ride-sourcing over transit, which depended on its quality and the city they operate. Our results from Fig. 2.15C presents a complementary perspective for the previous. Ride-sourcing's positive effects over transit may be a direct reflection of the negative externalities of ride-sourcing fleets over congestion. Hence, in areas where ride-sourcing appears to foster transit ridership, one should look for traffic congestion worsening and its causes.

Using the same assumption from Zha et al., 2018, we consider that ride-sourcing drivers

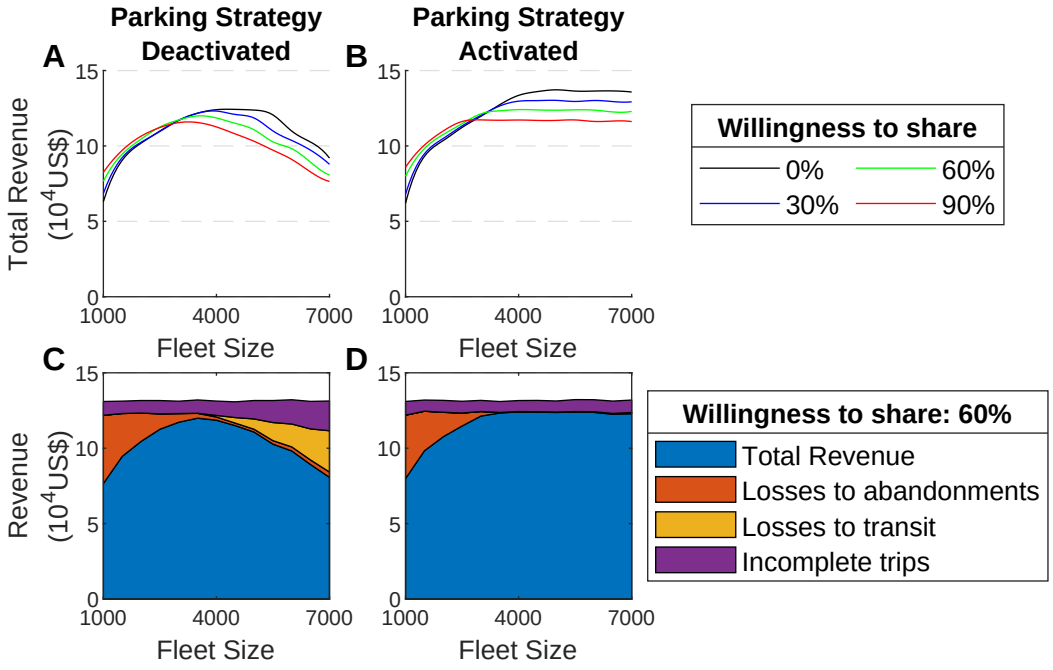


Figure 2.15: The system's revenues peak for particular fleet sizes and decrease due to congestion losses yielding higher revenues for lower willingness to share. (A and B) Total revenue of the system for increasing fleet sizes. (C and D) Potential revenues and summary of losses for scenarios with willingness to share 60%.

decide, about offering rides, daily. In general, drivers compete for passengers. In case the number of drivers is too large, very few will have satisfactory revenues that day, and most of the drivers will not drive the next days. Fig. 2.16 explores drivers' revenues after a 'day' (3 hours) of work. The revenue for a driver is 75% of what the passenger paid (the remaining 25% is the commission of the TNC). We set an illustrative minimum wage of US$14.50, which drivers consider, after a workday, in their decision to remain working in the next days. The minimum wage is for illustration purposes only; we do not claim it to be realistic. The $14.50 is a base value, which is equivalent to the American minimum wage (US$7.25 per hour) for a two-hour job. In other words, that is a case where drivers expect to be busy for about 2 hours. Willingness to share created higher variability in drivers' revenues and lower averages. In general, the average revenue ranges from US$21.24 and US$23.15. There were around 700 more drivers with revenues higher than the illustration when willingness to share is 0%, compared to it at 90%. The parking strategy, however, increased the average revenue by US$2.00 for a willingness to share of 0%. At the same time, only 2'896 drivers had higher revenues than the illustration threshold, 600 fewer drivers than the scenario where it is deactivated. Such findings point that the last-come-first-served basis for the queue inside the parking lot causes higher variability and, therefore, fewer drivers with satisfactory revenues. Note that many drivers had no revenue at all for a higher willingness to share.

Finally, in Fig. 2.16C, we can see that the system has plenty of scenarios for which the number

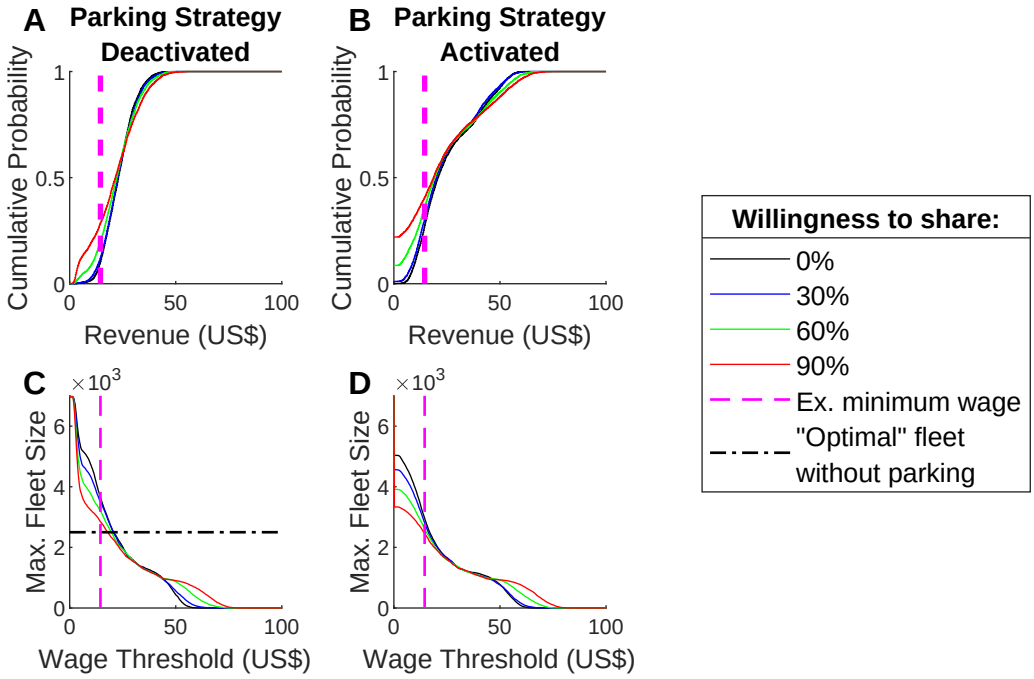


Figure 2.16: Analysis over drivers' revenue after one 'day' of work. (A and B) Cumulative probability of drivers' revenue for a fleet size of 4000 RSVs. (C and D) Maximum number of drivers with revenues higher than the wage threshold, i.e., drivers' sensitivity to wage thresholds. "Optimal" fleet stands for the fleet the minimizes the average journey duration in Fig. 2.6C.

of drivers with adequate revenues (the illustration wage threshold) is larger than the fleet size that minimizes journey durations (from Fig. 2.6C). There can be 1'200 more drivers than at these minimums. Furthermore, this number can increase even more if drivers' value of time (e.g., minimum wage threshold / wage reservation) is smaller. The previous indicates that more vulnerable areas may have more active drivers and more traffic issues (given that people still have access to private cars). Differently, the parking strategy made it less attractive to drivers, since only a few will have an acceptable wage (Fig. 2.16D). Note that it happens because of the queuing policy in the parking strategy, which is a *last-in-first-out* basis. It creates a situation in which few drivers to obtain higher revenues, while between 20% and 40% of driver in Fig. 2.16B obtain revenues lower than the minimum wage threshold; whereas, without a parking strategy, only 5% to 20% of drivers obtain revenues lower than the threshold (Fig. 2.16A). It is interesting to note that the number of drivers with satisfactory revenues becomes less sensitive to the starting fleet size. For all the previous, a policy suggestion would be to limit the number of simultaneous active drivers a TNC can have in the city.

Tirachini and Gomez-Lobo, 2019 analyzed for static conditions without congestion how incentives of the company and pricing strategies can influence the number of drivers that register for TNC services. Analyzing this type of equilibrium game for similar settings as our problem (with congestion dynamics, empty kilometer traveled) can reach additional

interesting insights. This should be a research priority.

Finally, results throughout the paper indicate the 'Wild Goose Chase' effect (Castillo et al., 2018; Xu, Yin, & Ye, 2020) for small fleet sizes, decreasing drivers revenues (Fig. 2.16) and a high number of drivers moving to pick-up passengers, compared to the number of drivers delivering them (Fig. 2.8 and Movie S1). Further investigation is needed in this direction.

2.4 Summary

In this chapter, we experimentally analyze the efficiency of TNCs using taxi trip data from a Chinese megacity and an agent-based simulation with a trip-based MFD model for determining the speed. We investigate the effect of expanding fleet sizes for TNCs, passengers' inclination towards sharing rides, and strategies to alleviate urban congestion. We observe that, although a larger fleet size reduces waiting time, it also intensifies congestion, which, in turn, prolongs the total travel time. Such congestion effect is so significant that it is nearly insensitive to passengers' willingness to share and flexible supply. Our findings also show that, even if drivers quit the system in case of low profit, the system can converge in a fleet size, which still causes noticeable congestion. Finally, parking management strategies can prevent idle vehicles from cruising without assigned passengers, mitigating the negative impacts of ride-sourcing over congestion, and improving the service quality.

3 A dynamic multi-region MFD model for ride-sourcing with ridesplitting

This chapter is based on the paper:

- Beojone, C. V., & Geroliminis, N. (2023a). A dynamic multi-region mfd model for ride-sourcing systems with ridesplitting. *Transportation Research Part B: Methodological, (Under Review)*

For the most recent version of this work, the reader can refer to the online version in the following URL: https://arxiv.org/abs/2211.14560.

3.1 Introduction

Ride-sourcing operators might not have a direct interest in congestion (see, for example, Chapter 2), but a dynamic model that captures congestion can be valuable for various operational decisions. While we developed a discrete-event simulation in Chapter 2 which allows for detailed analysis of the service and its interactions within a congested urban environment, it is not suited for fast near-future predictions (such as 30 minutes ahead in time), which makes it difficult to implement in real-time applications for repositioning or pricing optimization requiring testing of multiple potential solutions in a short period. Herein, we develop an MFD-based model representing ride-sourcing services and background traffic in a macroscopic multi-region urban network. The modeled ride-sourcing service offers ride-hailing (single rides) and ridesplitting (shared rides). Model states describe drivers on their ongoing activities and regions. We evaluated the proposed model by comparing the errors of the proposed model with benchmarks from the literature utilizing a detailed agent/trip-based simulator developed with real data from the the central business district of Shenzhen, China. Additionally, a sensitivity analysis investigated the generability of the model to assess the performance of multi-region traffic systems to several service parameters, such as fleet size, willingness to share or waiting time tolerance.

Following the motivation and challenges regarding the process of forecasting near-future

ride-sourcing and traffic conditions, which is given in Section 1.2.2 of Chapter 1, the remainder of the chapter has the following structure. Section 3.2 presents the general modeling framework of the proposed M-model for ride-sourcing with ridesplitting and its particular aspects applied to the modeled service operation. Section 3.3 evaluates the sensitiveness to service parameters and provide a few managerial insights. Section 3.4 depicts the numerical results of the proposed model directly compared to the plant and benchmark models and, finally, its robustness to noises in the input data.

3.2 General model framework

In ride-sourcing operations, a TNC uses a platform to centralize trip requests for services, such as ride-hailing and ridesplitting, from incoming passengers and to manage drivers who use their vehicles to profit from offering chauffeured rides (Rayle et al., 2016). In ride-hailing, single passengers (or a group of related travelers traveling together) request a ride in real-time and the operator tries to assign this trip to nearby drivers who should pick up the passenger and drive to a single destination directly. Ridesplitting allows multiple unrelated passengers to split rides if their routes overlap. Therefore, differently from ride-hailing, a detour can occur for at least one of the passengers served by the assigned driver. Usually, the only additional constraint concerns the added delay/detour to passengers.

Bringing a driver and a passenger together requires a matching process, usually focused on minimizing passengers' waiting times. If travelers wait too long, they might abandon the trip and use another mode of transport. For instance, TNCs, such as Uber, try to assign the closest vehicle to a new trip request on a *first-come-first-served* (FCFS) basis (Hanna et al., 2016). In this paper we make the following assumptions for the matching process of passengers to driver. Firstly, to maximize the chances of serving ridesplitting requests, the operator can consider interrupting ongoing trips with one passenger so that the vehicle changes its path to deliver both passengers. That means the operator makes real-time decisions and does not match passengers beforehand nor plans for interruptions. Secondly, we limit ridesplitting services to at most two simultaneous passengers per vehicle. Different matching strategies exist in the literature for example, perfect in advance knowledge (Santi et al., 2014), batch matching (Alonso-Mora et al., 2017) and others (Berbeglia et al., 2010; Jung et al., 2016; Ramezani & Nourinejad, 2018). In this case while different matching processes will require to revisit some aspects of the model, the dynamic framework can still be applied with little extra effort.

The proposed model describes ride-sourcing drivers based on their service assignments, following the provided operation description. The designed framework also incorporated urban traffic dynamics, tracking private vehicle activities, which formed the majority of background traffic. Therefore, all drivers in the model fit one of the activities below.

- Idle (I): a ride-sourcing vehicle with no assignments. It is vacant and available for any new passengers.

- Ride-hailing (RH): a ride-sourcing vehicle with an assigned ride-hailing passenger. Or the driver is moving to a pick-up location (Origin) or carrying a ride-hailing passenger towards the destination.
- Single ridesplitting ($S1$): a ride-sourcing vehicle with a single assigned ridesplitting passenger. Or the driver is moving to an origin or carrying a ridesplitting passenger towards the destination. A second ridesplitting assignment can interrupt this service.
- Shared ridesplitting ($S2$): a ride-sourcing vehicle that has two ridesplitting passengers assigned. Or the driver is moving to one of the pick-up locations or carrying two ridesplitting passengers towards one of their destinations.
- Private vehicle (PV): a private vehicle (outside the ride-sourcing service) traveling to the destination.

Note that we merged pick-up and delivery activities, using a single activity that indicates the existence of an assignment to simplify the state space and make the model less susceptible to noises in measurements. For instance, a driver picking up a ride-hailing passenger is simply in a ride-hailing activity, similarly to a driver who is delivering a ride-hailing passenger is also in a ride-hailing activity. The same occurs in the case of ridesplitting activities, and the difference between $S1$ and $S2$ activities is the number of passengers assigned to the driver. For instance, a driver with one in-vehicle passenger and picking up the second one is in a $S2$ activity because it has two assigned passengers (the one inside the vehicle and the one to be picked up).

3.2.1 Macroscopic model dynamics and mass conservation equations

Besides service-related transitions, drivers will experience different traffic situations while they move on the road network, depending on their current region. MFD models can describe dynamic state evolution for urban networks partitioned into multiple homogeneously congested regions. The proposed model uses MFD dynamics to compute the flows of ride-sourcing and private vehicles in a macroscopic urban network. For illustration, it is composed of a set $\mathscr{R}$ with R heterogeneous regions, i.e., $\mathscr{R} = \{1, 2, ..., R\}$, each with a well-defined speed-MFD expressing regional speeds as a function of accumulation $v_o(t) = V_o(n_o(t))$. Therefore, traffic congestion and average speeds are functions of a Speed-MFD, of which the accumulation is the sum of private and ride-sourcing vehicles. We can scale the function to represent homogeneously congested portions of the area, analogously to Ni and Cassidy, 2020.

We developed a multi-region M-model which only focuses on vehicular traffic to represent private and ride-sourcing vehicles in different states, which are described based on their activities $K \in \mathbb{A}$ (where $\mathbb{A} = \{I, RH, S1, S2, PV\}$ is the set of activities) and the current and destination regions $od \in \mathscr{R}^2$, summarized into the notation K_{od}. Note that the set of activities $\mathbb{A}$ includes all previously mentioned activities: idle (I), ride-hailing (RH), single ridesplitting ($S1$), shared ridesplitting ($S2$), and private vehicle (PV). Two sets of conservation equations

describe the dynamics of each state (Equations [3.1]–[3.4]). The first one computes the evolution of the number of vehicles, and the second one, of the total remaining distance. Idle vehicles are the only exception without remaining distance to be estimated because they have no assignments to complete. Differently than the classical MFD approach, where a vehicle that starts a trip will finalize it with a specific trip length, ridesplitting services contain this additional complexity because of the interruptions. While a vehicle in an activity $S1$ contributes in the remaining distance of this state with a pre-determined trip length, when a second passenger is assigned the state changes from $S1$ to $S2$ without completing the $S1$ trip, creating an inconsistency in the classical MFD framework (a trip that starts needs to complete its assigned trip length). That is the reason for $S1_{od}$ being a special case with their particular dynamics accounting for such trip interruptions. Thus, we only need Equation [3.1] to depict idle drivers' dynamics. In summary, the number of states can be computed as $|\mathscr{K}_{od}| = |\mathscr{R}| + (|\mathbb{A}| - 1) \cdot |\mathscr{R}|^2$, where $\mathscr{K}_{od}$ is the set of all states.

$$\dot{n}^K_{od}(t) = \text{Inflow} - \text{Outflow} \qquad K \in \mathbb{A} \setminus S1 \tag{3.1}$$

$$\dot{M}^K_{od}(t) = \text{Inflow} \cdot \text{Trip length} - n^K_{od}(t)\, v_o(t) \qquad K \in \mathbb{A} \setminus \{I, S1\} \tag{3.2}$$

$$\dot{n}^{S1}_{od}(t) = \text{Inflow} - \text{Outflow} - \text{Interruption} \tag{3.3}$$

$$\dot{M}^{S1}_{od}(t) = \text{Inflow} \cdot \text{Trip length} - n^{S1}_{od}(t)\, v_o(t) - \text{Interruption} \cdot \text{Remaining distance} \tag{3.4}$$

where 'Inflow', 'Outflow' and 'Trip length' are defined for each state in Table 3.1. Besides trip lengths, described using the respective $L^K_{od}(t)$, some of the main components of the dynamics $O^K_{od}(t)$ and $O^K_{ohd}(t)$ are trip completion and transfer flow rates, respectively; where o, h and d represent the current, the next and the final region of drivers' path. On the other hand, $\bar{\lambda}^P_{od}(t)$ (where $P \in \{RH, S1\}$) and $\bar{\lambda}^{S2}_{ohd}(t)$ summarize traveler entering processes assigned to idle drivers or shared ridesplitting rides, respectively; and the regions o, h, and d refer to the origin, the intermediate stop of a shared request, and the final destination of an arriving request, respectively.[I] Private vehicles have their own arrival/enter process for travelers depicted by the value $\bar{\lambda}^{PV}_{od}(t)$. Equations [3.1] and [3.2] did not include vehicles in $S1$ activities due to possible interruptions. These interfere with the total remaining distance, meaning that part of the production – total distance traveled per time unit – does not directly convert into trip completion or transfer flows. A general state-space framework is illustrated in Figure 3.1 focusing on the transitions inside one individual region and its interactions with the neighboring ones, where any region k is preceding region o, which is preceding any region l on drivers' path to region d (or region o for those trips ending there).

Note that there is no endogenous trip completion for idle I_o drivers since they do not have any assignments to complete; instead, they cruise for passengers, and they exit their current state through passenger entering processes. While one could compute *a posteriori* what is the

[I] Note that the willingness to share, as explained in Chapter 2, will separate the passenger arriving process into the different service options.

Table 3.1: Summary of dynamic flows in each state (notation described in table 4).

State	Nb. of vehicles	Rem. distance	Inflow	Outflow	Trip length
I_o	$n_o^I(t)$	–	$O_{oo}^{RH}(t) + O_{oo}^{S1}(t)$	$\sum_{d\in\mathscr{R}} \bar{\lambda}_{od}^{RH}(t) + \bar{\lambda}_{od}^{S1}(t)$	–
RH_{od}	$n_{od}^{RH}(t)$	$M_{od}^{RH}(t)$	$\bar{\lambda}_{od}^{RH}(t) + \sum_{i\in\mathscr{R}_o} O_{iod}^{RH}(t)$	$O_{od}^{RH}(t)$	$L_{od}^{RH}(t)$
$S1_{od}$	$n_{od}^{S1}(t)$	$M_{od}^{S1}(t)$	$\bar{\lambda}_{od}^{S1}(t) + \sum_{i\in\mathscr{R}_o} O_{iod}^{S1}(t) + O_{ood}^{S2}(t)$	$O_{od}^{S1}(t)$	$L_{od}^{S1}(t)$
$S2_{od}$	$n_{od}^{S2}(t)$	$M_{od}^{S2}(t)$	Interruptions $+ \sum_{i\in\mathscr{R}_o} O_{iod}^{S2}(t)$	$O_{od}^{S2}(t)$	$L_{od}^{S2}(t)$
PV_{od}	$n_{od}^{PV}(t)$	$M_{od}^{PV}(t)$	$\bar{\lambda}_{od}^{PV}(t) + \sum_{i\in\mathscr{R}_o} O_{iod}^{PV}(t)$	$O_{od}^{PV}(t)$	$L_{od}^{PV}(t)$

average distance traveled for vehicles in state I_o, this is not defined in the classical way as in MFD models because it is state-specific and varies over time.

In Table 3.1, inflows and outflows of state RH_{od} illustrate that drivers do not deliver their passengers before entering the destination region. Every new ride-hailing assignment adds its average trip length to the remaining distance. The model assumes, without loss of generality, a single average trip length for all input flows.[II] Later, in Section 3.2.4 the calculation of every trip length is detailed.

Most inflows and outflows for ridesplitting are naturally compatible with those for ride-hailing. However, the traveler entering process has a double role. While a portion serves as inflow for state $S1_{od}$, the remaining interrupts a current service, as an additional passenger is entering the vehicle. Traditionally, in MFD-based models, vehicles must always complete the started trips. Interruptions violate such an assumption. Therefore, Equations [3.3] and [3.4] relax this assumption for state $S1_{od}$.

In Equation [3.5], we separate the interruptions depending on the trip scheme resulting from the matching process, namely *last-in-first-out* (LIFO) and *first-in-first-out* (FIFO), where the last and first refer to the passengers entering and exiting the vehicle. Note that $\bar{\lambda}_{ohd}^{S2}(t)$ and $\bar{\lambda}_{odh}^{S2}(t)$ represent the rate of ridesplitting requests assigned into a LIFO and a FIFO shared ridesplitting trip-schemes, respectively.[III] The first part, where we have $\sum_{h\in\mathscr{R}} \bar{\lambda}_{ohd}^{S2}(t)$, refers to cases where the destination of the new traveler lies in one of the possible regions on the path of the initial trip, being, thus, delivered earlier than the initial passenger (LIFO trip scheme). The second part, where we have $\sum_{h\in\mathscr{R}} \bar{\lambda}_{odh}^{S2}(t)$, refers to cases where the destination of the new traveler is farther than the initial one, being, thus, delivered after the initial passenger

[II] To relax this assumption, one must use individual trip lengths for each of the entering possibilities in the state. Then, the computation of $O_{od}^{K}(t)$ can later use a weighted average (based on the input rates) of these trip lengths.

[III] Recall that requests are classified as ridesplitting at the moment the passenger hires the service, where they become classified as "willing to share".

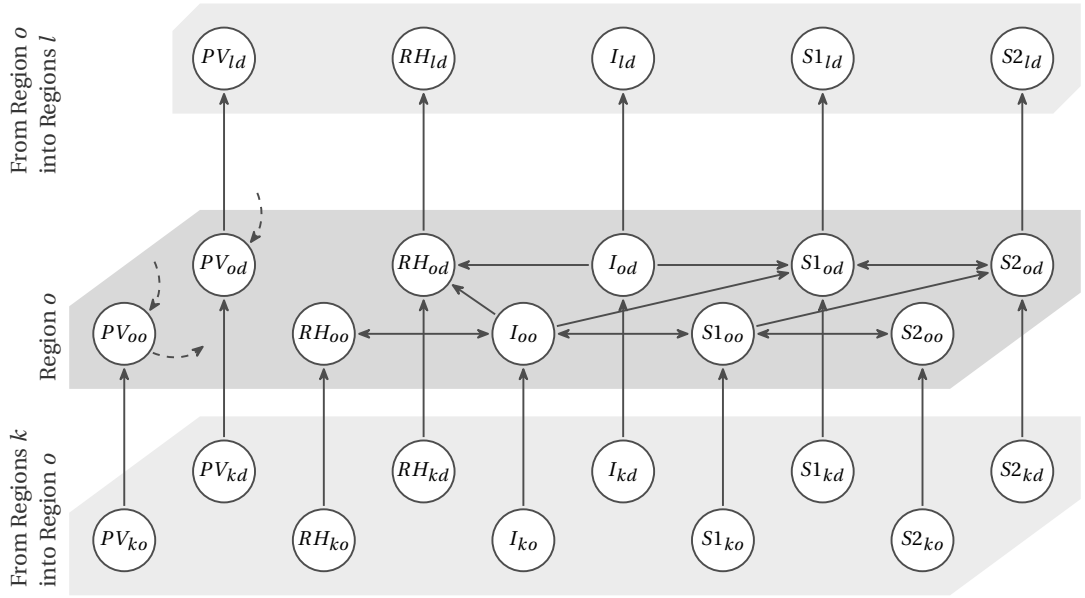


Figure 3.1: General state transition structure.

(FIFO trip scheme). In LIFO trip schemes, drivers in state $S1_{od}$ will enter a state $S2_{od}$ for an incoming oh request. In FIFO trip schemes, drivers in state $S1_{od}$ will enter a state $S2_{oh}$ for an incoming oh request. Note that there must be a similarity between the od and oh so that $\bar{\lambda}^{S2}_{ohd}(t) > 0$ or $\bar{\lambda}^{S2}_{odh}(t) > 0$. We describe in detail how to estimate these values later in Section 3.2.3, for now it stands for the general framework of the proposed model.

$$\text{Interruptions} = \underbrace{\sum_{h \in \mathscr{R}} \bar{\lambda}^{S2,\text{LIFO}}_{ohd}(t)}_{\text{LIFO trip scheme}} + \underbrace{\sum_{h \in \mathscr{R} \setminus \{d\}} \bar{\lambda}^{S2,\text{FIFO}}_{odh}(t)}_{\text{FIFO trip scheme}} \tag{3.5}$$

We only accounted for $\bar{\lambda}^{S2}_{odd}(t)$ in the LIFO trip scheme to avoid double counting.

The last element of Equation [3.4] is the 'Remaining distance' $L^{*}_{S1_{od}}(t)$ at the moment of the second assignment (interruption). It illustrates the effect of not completing a trip as initially planned, while the 'Interruption' illustrates the process.

Recall that the entering flow of passengers in activity $S2$ is equivalent to the interruptions in activity $S1$. Drivers in $S2_{od}$ carry two passengers with possibly different destination regions but with a similar regional path. Therefore, these drivers might drop a passenger before reaching d. This results from the possible trip schemes in shared rides. The trip length $L^{S2}_{od}(t)$ has no relation to the remaining distance $L^{*}_{S1_{od}}(t)$.

Finally, throughout the paper, we refer to trip length as the space a driver travels in a region

(focus of the traffic model), not the whole distance that a driver covers from the assignment until the drop-off area. Although related, they represent different aspects of the ride-sourcing operation in a traffic model. While the first one relates to drivers' network movements, the second one relates to individual service requests. We should mention that even though some regional trip lengths are the same for different services/activities, the total trip length from the time of a vehicle assigned to the first passenger, until completing the trip varies across time and sequence of activities.

3.2.2 Trip completion rates, transfer flows

Once assigned, the driver enters a busy state and starts one of the outflow processes of trip completion or transfer. In terms of notation, the difference between them is the state description, where state K_{od} completes a trip if $o = d$ or transfers if $o \neq d$. Vehicles outside their region of destination must first transfer along the trip path and then complete the trip. The proposed M-model computes outflows using Equation [3.6]. Recall that I_o drivers' outflow is the passenger entering process.

$$O^K_{od}(t) = \frac{n^K_{od}(t)\, v_o(t)}{L^K_{od}(t)} \left(1 + \alpha \left(\frac{M^K_{od}(t)}{n^K_{od}(t) L^*_{K_{od}}} - 1\right)\right), \qquad K \in \mathbb{A}\backslash\{I, S1\}, \text{ and } \{o,d\} \in \mathscr{R}^2 \tag{3.6}$$

where $O^K_{od}(t)$ is the instantaneous trip completion/transfer rate for vehicles in state K_{od}; α is a model parameter expressing the sensitivity of outflow to variations in the remaining distance $M^K_{od}(t)$. L^K_{od} is the average trip length. $L^*_{K_{od}}$ is the steady-state average remaining distance in state K_{od} until exiting the current region. It can be computed as a function of the average trip length $L^K_{od}(t)$ and its standard deviation $\sigma^K_{od}(t)$, i.e., $L^*_{K_{od}}(t) = \left(L^K_{od}(t)^2 + \sigma^K_{od}(t)^2\right)\left(2L^K_{od}(t)\right)^{-1}$.

Recall that the interruption of an ongoing $S1_{od}$ delivery for a new shared ride leaves a $L^*_{S1_{od}}(t)$ distance to the destination or next region uncovered, meaning that part of the production of these vehicles does not convert into trip completion or transfer rates. We use mass conservation Equations [3.3] and [3.4] at steady-state to estimate an approximation of $O^{S1}_{od}(t)$. The idea in Equations [3.7]–[3.8] is to isolate the inflow to combine both equations and obtain the estimate in Equation [3.9].

$$\dot{n}^{S1}_{od}(t) = \text{Inflow} - O^{S1}_{od}(t) - \text{Interruption} = 0 \Rightarrow$$
$$\Rightarrow \text{Inflow} = O^{S1}_{od}(t) + \text{Interruption} \tag{3.7}$$
$$\dot{M}^{S1}_{od}(t) = \text{Inflow} \cdot L^{S1}_{od}(t) - n^{S1}_{od}(t)\, v_o(t) - \text{Interruption} \cdot L^{*}_{S1_{od}}(t) = 0 \Rightarrow$$
$$\Rightarrow \text{Inflow} = \frac{n^{S1}_{od}(t)\, v_o(t) + \text{Interruption} \cdot L^{*}_{S1_{od}}(t)}{L^{S1}_{od}(t)} \tag{3.8}$$
$$\begin{aligned} O^{S1}_{od}(t) &= \frac{n^{S1}_{od}(t)\, v_o(t) + \text{Interruption} \cdot L^{*}_{S1_{od}}(t)}{L^{S1}_{od}(t)} - \text{Interruption} \\ &= \frac{n^{S1}_{od}(t)\, v_o(t)}{L^{S1}_{od}(t)} - \left(1 - \frac{L^{*}_{S1_{od}}(t)}{L^{S1}_{od}(t)}\right) \cdot \text{Interruption} \\ &= \hat{O}^{S1}_{od}(t) - \left(1 - \frac{L^{*}_{S1_{od}}(t)}{L^{S1}_{od}(t)}\right) \cdot \text{Interruption} \end{aligned} \tag{3.9}$$

where $L^{S1}_{od}(t)$ is the trip length of newly assigned single ridesplitting trips; $\hat{O}^{S1}_{od}(t)$ is an estimator of the trip completion rate (or transfer flow) without interruptions; and the second term of the result is the amount of $\hat{O}^{S1}_{od}(t)$ to be discounted due to interruptions. Finally, one can estimate $\hat{O}^{S1}_{od}(t)$ using the M-model approximation from Equation [3.6].

Drivers may drive through different routes between their current and destination regions. The outflow $O^{K}_{ohd}(t)$ is the transfer rate from a current region o through the immediate next one h, as illustrated in Equation [3.10]. Therefore, $\theta_{ohd} \in [0,1]$ distributes transfer flows over its neighboring regions such that the equality $\sum_{h \in \mathscr{R}_o} \theta_{ohd} = 1$ holds. The internal trip completion rate is computed directly from Equation [3.6], where $o = d$.

$$O^{K}_{ohd}(t) = \theta_{ohd} \cdot O^{K}_{od}(t) \qquad K \in \mathbb{A} \setminus \{S2\} \tag{3.10}$$

Drivers in $S2_{od}$ are an exception to the previous because they may deliver one of their passengers in a region before the last destination. The process for dropping one of the passengers in the current region precedes transfer. Hence, part of the drivers will transfer (Equation [3.11]), while others will return to state $S1_{od}$ before transferring (Equation [3.12]).

$$O^{S2}_{ohd}(t) = \theta_{ohd} \cdot (1 - \vartheta_{ood}(t)) \cdot O^{S2}_{od}(t) \qquad h \in \mathscr{R}_o \tag{3.11}$$
$$O^{S2}_{ood}(t) = \vartheta_{ood}(t) \cdot O^{S2}_{od}(t) \tag{3.12}$$

where $\vartheta_{ood}(t)$ becomes the fraction of shared trips passing through o that will deliver a

passenger before continuing to d.

Regional trip information is essential for MFD-based models, and, in the case of ride-sourcing, it results from the passenger-driver matching process. Ride-sourcing drivers pick up and then deliver their assigned passengers and Equation [3.13] breaks trip lengths into both activities. Recalling that, trip lengths in the proposed model are the distances traveled in a region, not the ones for whole assignments. In this paper we assume that the values for θ_{ohd} are exogenously defined, while there are approaches in the MFD literature to integrate through an aggregated assignment process, which is beyond the scope of this work (see for example Yildirimoglu and Geroliminis, 2014). The reader interested in obtaining regional trip lengths and their distribution for MFD-based models is invited to read Batista et al., 2019 and the references therein.

$$L^{K}_{od}(t) = L^{\text{pick}}_{K_{od}}(t) + L^{\text{drop}}_{K_{od}}(t) \tag{3.13}$$

3.2.3 Drivers movements and passenger-driver matching process

Matching passengers with available vehicles is at the core of ride-sourcing operations. It defines if an arriving passenger enters the system or leaves unserved. Replicating it in a dynamic model requires translating such a microscopic activity into a macroscopic scheme. We intended to identify a parsimonious way to integrate matching in an aggregated model.

Most efforts of replicating this process into a dynamic model use Cobb-Douglas matching friction function. Examples of such include Ramezani and Nourinejad, 2018, Xu, Yin, Chao, et al., 2020, and Nourinejad and Ramezani, 2020. However, it assumes that passengers wait for an amount of time and form a batch to assign them to drivers, such that it balances demand and the supply of drivers. None of the previous studies handled large-scale ride-sourcing with a ridesplitting option with an FCFS assignment.

As a consequence of the FCFS assignment, we can formulate a loss probability function in response to endogenous variables such as the available fleet size $n^{av}_{od}(t)$, the regional average speed $v_o(t)$, the waiting time tolerance ω, and the ratio of Idle-Busy drivers ρ^{s}_{od}.[IV] It constitutes an operational result for a given demand profile and response to service quality requirements. The function should yield a few properties that will later ensure non-negativity to drivers' numbers and steer the consequences of the matching process, including the thickness to demand. Firstly, $pl^{s}_{od} \in (0,1]$ for $n^{av}_{od}, v_o, \omega, \rho^{s} \geq 0$. Secondly, if any of the parameters approaches

[IV] We must highlight that this approximation may be revised in the case of the matching process batch passengers before the assignment and cases where the supply of drivers is too small. For instance, if one desires to batch passengers, instead of loss probabilities, one can create a state of waiting passengers in each region, and use a Cobb-Douglas function or an input-output diagram to compute the number of matches after batching some passengers and drivers (as already seen in the literature before). Furthermore, to integrate multiple degrees of patience, one can group passenger arrival processes according to their waiting time tolerances.

0, then pl^s_{od} approaches 1. Thirdly, all partial derivatives are negative, i.e., all parameters decrease the chances of losing incoming requests. With these assumptions, a Cobb-Douglas type function fits the negative log probability to ensure the previously mentioned desired properties. Equation [3.14] depicts the formulation.

$$pl^s_{od}(n^{av}_{od}, v_o, \omega, \rho^s_{od}) = \exp\left(-\gamma_0 \cdot (n^{av}_{od})^{\gamma_1} \cdot (v_o)^{\gamma_2} \cdot \omega^{\gamma_3} \cdot (\rho^s_{od})^{\gamma_4}\right) \tag{3.14}$$

All parameters $\gamma_q, q \in \{0,1,2,3,4\}$ must be positive to ensure negative partial derivatives. One must acknowledge that γ_0 represents the coverage of a vehicle, while γ_1, γ_2, γ_3, γ_4 indicate the coverage sensitivity to each endogenous variable. Furthermore, the these parameters are not dependent on the service option. However, endogenous variables can vary between service options depending on particular aspects of the modeled system. We have to note that these curves might not be universal and depend on the matching policy.

Note that the computation of pl^s_{od} considered all available drivers $n^{av}_{od}(t)$ because one cannot identify which drivers would be capable of serving the arriving request before the assignment. Therefore, if we account only for drivers that would comply with all the constraints, we would be changing the sample space, in an example of the "Monty Hall" problem.

A Monte Carlo simulation followed by a linear regression model evaluated the parameters $\gamma_q, q \in \{0,1,2,3,4\}$ of Equation [3.14] to compute the loss of incoming requests after checking their feasibility constraints (waiting time and/or detour). Appendix B details the construction and algorithm of the Monte Carlo simulation, while Appendix C summarizes how Equation [3.14] was linearized to construct a linear regression model with the results of the Monte Carlo simulation. For instance, the linear regression of a single region experiment with the entire road network obtained a $R^2 = 0.96$.

Since ride-sourcing cannot serve all arriving customers, exogenous arrival rates enter the ride-sourcing restrained by the respective loss probabilities $pl^s_{od}(t)$ (simpler notation of Equation [3.14]) in Equation [3.15]. We assume that lost customers use private vehicles, so we penalize the congestion and maintain the total number of trips. Equation [3.16] adds these lost requests to the private vehicle demand.

$$\bar{\lambda}^s_{od}(t) = \left(1 - pl^s_{od}(t)\right)\lambda^s_{od}(t), \qquad \text{where, } s \in \{H, S\} \tag{3.15}$$

$$\bar{\lambda}^{PV}_{od}(t) = \lambda^{PV}_{od}(t) + \sum_{s \in \{H,S\}} pl^s_{od}(t)\lambda^s_{od}(t) \tag{3.16}$$

where $\lambda^s_{od}(t)$ and $\lambda^{PV}_{od}(t)$ are the arrival rate of travelers for one of the ride-sourcing services (ride-hailing H or ridesplitting S) and private vehicles, respectively. Then, $\bar{\lambda}^s_{od}(t)$ and $\bar{\lambda}^{PV}_{od}(t)$

are the entrance rate of these travelers (we distinguish between arrival and entrance).

The proposed model assumes a similar geographical distribution among all available drivers for a service. Therefore, Equation [3.17] can endogenously compute the proportion of entering passengers assigned to idle drivers, $\rho^s_{od}(t)$, based on the instantaneous number of drivers; while the probability of assigning it to a busy vehicle is $1-\rho^s_{od}(t)$.[V,VI] The previous assumes that available vehicles and arriving passengers must be in the same region. Such an assumption is reasonable if the number of assigned drivers across the region limits is negligible. The number of available drivers for a shared ridesplitting ride differs, depending on the evaluated trip-scheme. Therefore, Equations [3.18] and [3.19] counts the number of available drivers for LIFO and FIFO trip-schemes, respectively.

$$\rho^s_{od}(t) = \frac{n^I_o(t)}{n^{av}_{od}(t)} = \begin{cases} \dfrac{n^I_o(t)}{n^I_o(t)} = 1, & \text{if } s = H \\ \dfrac{n^I_o(t)}{n^I_o(t) + n^{av,\text{LIFO}}_{od}(t) + n^{av,\text{FIFO}}_{od}(t)}, & \text{if } s = S \end{cases} \tag{3.17}$$

$$n^{av,\text{LIFO}}_{od}(t) = \sum_{h \in \mathscr{R}} \beta^d_{oh} n^{S1}_{oh}(t) \tag{3.18}$$

$$n^{av,\text{FIFO}}_{od}(t) = \sum_{h \in \mathscr{R} \setminus \{d\}} \beta^h_{od} n^{S1}_{oh}(t) \tag{3.19}$$

where β^h_{od} (β^d_{oh}) represents the ratio of od (oh) trips that will pass through region h (d).

The computation of the loss probability $pl^s_{od}(t)$ considers that unacceptable detours will restrain some of the demand from entering the service (see the Appendix B for details). Therefore, one should not limit available drivers $n^{av}_{od}(t)$ to those complying with all constraints (wait and detour), under the penalty of accounting twice for the same effects.

We can further divide the passenger entrance process according to the activity assigned to the driver. In the case of ride-hailing, Equation [3.20] confirms that entering passengers causes idle drivers to enter state RH_{od}. For ridesplitting, Equation [3.21] states the process of having drivers assigned to single requests. Equations [3.22] and [3.23] illustrate the assignment of drivers to shared ridesplitting requests in LIFO and FIFO trip schemes, respectively.

[V] One can extend the ratios $\rho^s_{od}(t)$ to larger passenger capacities if the assumptions remain the same for drivers' geographical distribution and dispatching policy.

[VI] Cases with distinct dispatching policies may require direct prioritization of certain vehicles (such as priority to vehicles at state $S1$), requiring adaptations to the computation of the $\rho^s_{od}(t)$ and $pl^s_{od}(t)$.

$$\bar{\lambda}^{RH}_{od}(t) = \rho^{H}_{od}(t)\bar{\lambda}^{H}_{od}(t) = \bar{\lambda}^{H}_{od}(t) \tag{3.20}$$

$$\bar{\lambda}^{S1}_{od}(t) = \rho^{S}_{od}(t)\bar{\lambda}^{S}_{od} \tag{3.21}$$

$$\bar{\lambda}^{S2,\text{LIFO}}_{ohd}(t) = \left(1-\rho^{S}_{oh}(t)\right)\cdot\frac{\beta^{h}_{od}n^{S1}_{od}(t)}{n^{av,\text{LIFO}}_{oh}(t)+n^{av,\text{FIFO}}_{oh}(t)}\bar{\lambda}^{S}_{oh}(t) \tag{3.22}$$

$$\bar{\lambda}^{S2,\text{FIFO}}_{odh}(t) = \left(1-\rho^{S}_{oh}(t)\right)\cdot\frac{\beta^{d}_{oh}n^{S1}_{od}(t)}{n^{av,\text{LIFO}}_{oh}(t)+n^{av,\text{FIFO}}_{oh}(t)}\bar{\lambda}^{S}_{oh}(t) \qquad h\neq d \tag{3.23}$$

In the assumed matching process, one of the passengers will be delivered first in shared ridesplitting. Equation [3.24] uses current demand information to identify the fraction of shared trips delivering a passenger in their current region, $\vartheta_{ood}(t)$.[VII]

$$\vartheta_{ood}(t) = \frac{\sum_{h\in\mathscr{R}}\bar{\lambda}^{S2}_{hod}(t)}{\sum_{h\in\mathscr{R}}\sum_{l\in\mathscr{R}}\beta^{o}_{hl}\beta^{l}_{od}\bar{\lambda}^{S2}_{hld}(t)} \tag{3.24}$$

where, $\sum_{h\in\mathscr{R}}\sum_{l\in\mathscr{R}}\beta^{o}_{hl}\beta^{l}_{od}\bar{\lambda}^{S2}_{hld}(t)$ indicates all demand for shared ridesplitting trips heading to region d that will pass through o (either for delivering a passenger, or just as a passage towards d) before delivering one of the passengers.

3.2.4 Trip length estimates

In assuming a FCFS matching process to the nearest available driver, pick-up trip lengths become analogous to the average minimal distance to the center of a circle. Daganzo, 2010 and Daganzo and Ouyang, 2019 derived the approximation in Equation [3.25], where the product between waiting time tolerance ω and instantaneous speed $v_o(t)$ determines the matching radius. The literature presents other similar results (K. Zhang et al., 2019; K. Zhang & Nie, 2019). From the loss probability, we approximate the number of matchable drivers (number of drivers that are available and comply with all matching constraints) as $(1-pl^{s}_{od}(t))n^{av}_{od}(t)$.

$$L^{\text{pick}}_{K_{od}}(t) \approx 0.63\sqrt{\frac{\omega\cdot v_o(t)}{(1-pl^{s}_{od}(t))\cdot n^{av}_{od}(t)}} \tag{3.25}$$

We averaged time invariant trip lengths for delivery, $L^{\text{drop}}_{K_{od}}$, based on historical data. The previous refers only to intra-regional trip lengths in the historical data, not the length of multi-

[VII] Tracking every delivery stop in the state notation would relax this memoryless assumption by deteriorating model's scalability.

region trips, which would vary according to the demand (Origin-Destination combination) and vehicle routing choices.

The model also uses the remaining distance $L^*_{K_{od}}$ in outflows (trip completion and transfer flows) and interruptions. As mentioned earlier, in steady-state $L^*_{K_{od}}(t) = \left(L^K_{od}(t)^2 + \sigma^K_{od}(t)^2\right) \cdot \left(2L^K_{od}(t)\right)^{-1}$. Demand changes may take the traffic system out of a steady-state condition, changing the actual value of $L^*_{K_{od}}$.

The coefficient of variation (σ/L) remained almost constant in historical data (computed from several simulations, with the data aggregated in time intervals of 3 minutes for each activity), ranging between 0.54 and 0.64 (depending on the experimental settings in the simulation, OD-pair and number of regions), which is far from a value of 1, justifying the choice of an M-model. Furthermore, given the nearly constant coefficients of variation, there is no need to compute $\sigma^K_{od}(t)$ separately, simplifying the computation of $L^*_{K_{od}}$.

3.3 Model's sensitivity analysis in a multi-region setting

A multi-region setting can provide insights for those interested in developing regulatory schemes envisioning better traffic conditions and general welfare. Nevertheless, the platform operator can also derive rapid forecasts and evaluate possible decisions and near-future consequences in the operation of the service. Therefore, in this section, we assume a three-region model described with the equations of Section 3.2 and we investigate the dynamic evolution of states for different fleet size, willingness to share (how the arriving demand $\lambda(t)$ is separated between ride-hailing $\lambda^H(t)$ and ridesplitting $\lambda^S(t)$) and waiting time tolerance with the parameters depicted in Table 3.2. Differently from Chapter 2, we separate the analyzed area into three regions to capture the model features related with a multi-regional setting (e.g. ridesplitting assignments and their matching constraint checks). We computed the data based on the demand data from Shenzhen (Bellocchi & Geroliminis, 2020). Other data include a constant coefficient of variation $\sigma/L = 0.650$ and the Speed-MFD, as depicted in Figure 3.2.

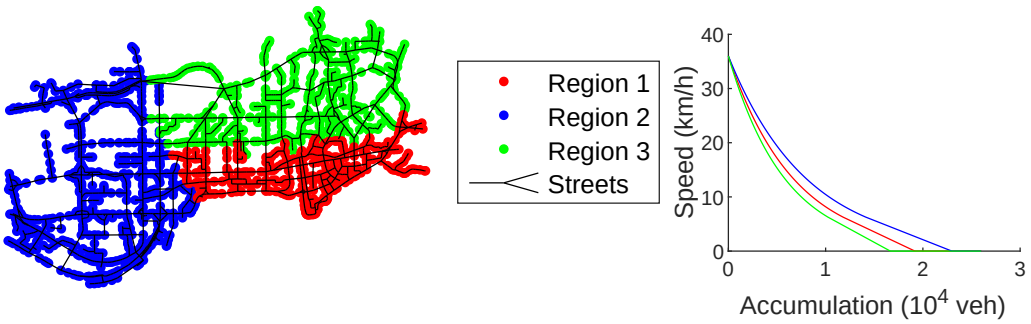


Figure 3.2: Shenzhen central business district separated in three regions and their respective Speed-MFDs used in the sensitivity analysis.

Table 3.2: Three-region model parameters.

OD-pair	Demand ratio	Transfer ratios (θ_{okd})			Passage ratio (β^k_{od})			$L^{\text{drop}}_{K_{od}}$ (km)		
		$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$	RH/PV	$S1$	$S2$
1-1	0.136	n/d	n/d	n/d	1	0	0	2.119	2.084	2.144
1-2	0.070	0	0.916	0.084	1	1	0.084	2.805	2.860	2.562
1-3	0.099	0	0.005	0.995	1	0.004	1	1.861	2.059	1.923
2-1	0.061	0.904	0	0.096	1	1	0.095	3.257	3.307	2.758
2-2	0.308	n/d	n/d	n/d	0	1	0	3.128	3.148	3.074
2-3	0.055	0.535	0	0.465	0.534	1	1	3.304	3.079	3.015
3-1	0.091	0.995	0.005	0	1	0.005	1	2.095	2.168	1.855
3-2	0.054	0.449	0.551	0	0.449	1	1	2.820	2.807	2.579
3-3	0.126	n/d	n/d	n/d	0	0	1	2.292	2.297	2.274

(n/d: non-defined)

We set the plant to reproduce dynamic traffic entering the hyper-congested regime and then returning to an uncongested state without reaching gridlock. Such a setting is supposed to generate a challenging scenario for the model evaluation since hyper-congested situations create conditions far from steady-state and hysteresis during the loading and unloading of the network. A Poisson process describes the arrivals in piece-wise constant rates during 3 hours of simulation such that there is a peak hour preceded and followed by low-demand hours. It generates a total of 40,000 trips per hour during the low-demand hours and 70,000 trips per hour during the peak hour, from which 85% are background traffic and 15% are ride-sourcing requests.

Firstly, passengers may have different tolerances and service preferences, while the platform can manage its service to influence the number of active drivers and passengers' service choices. Therefore, in Figure 3.3, we quantify the effects of passengers' willingness to share (i.e., the fraction of ride requests for ridesplitting), their waiting time tolerance ω, and the fleet size of active ride-sourcing drivers. As expected, increased fleet sizes and willingness to share decreased the fraction of lost requests. For instance, with a waiting time tolerance of 60 seconds, a fleet size of about 2100 drivers reaches the same 15% abandonment ratio as a fleet of 2800 drivers when the willingness to share increases from 25% to 100%. However, waiting time tolerance creates a different behavior. On the one hand, more patient passengers enlarge the coverage area for pick-up, increasing the chances for passenger-driver matching. On the other hand, it allows assigning drivers farther from their passengers, which keeps them busy for prolonged periods, decreasing their availability for incoming requests, in one consequence of the wild-goose chase effect (Castillo et al., 2018).

As a direct consequence of the same settings, they affect the average waiting time, a key performance indicator to attract and maintain customers in this service. In Figure 3.4, passengers' waiting time tolerance causes the most significant changes in waiting time, changing average values in orders of magnitude from less than 10 seconds to near 10 minutes for tolerances of 1 and 10 minutes, respectively. Note that these numbers are also affected by an aban-

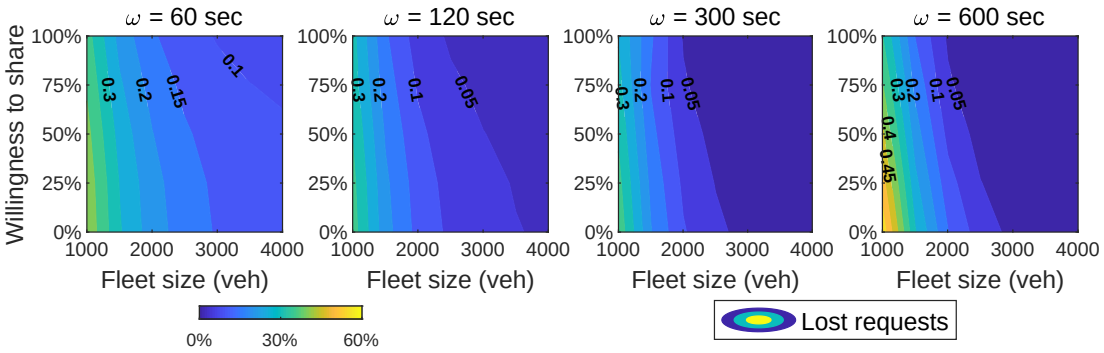


Figure 3.3: Summary of abandonment rates as a function of fleet size, willingness to share and waiting time tolerance (ω).

donment penalty, such that the average waiting time increases by the abandonment rate as Penalized waiting time = Waiting time $\cdot$ (1 + Abandonment) (Beojone & Geroliminis, 2021b). Enlarging the available fleet had higher impacts than passengers' willingness to share. Pairing both fleet size and willingness to share can achieve more efficient outcomes, such as keeping the same 20 seconds average waiting time by increasing willingness to share from 0% to 100% while having 1000 less active drivers (scenario with $\omega = 300$ seconds).

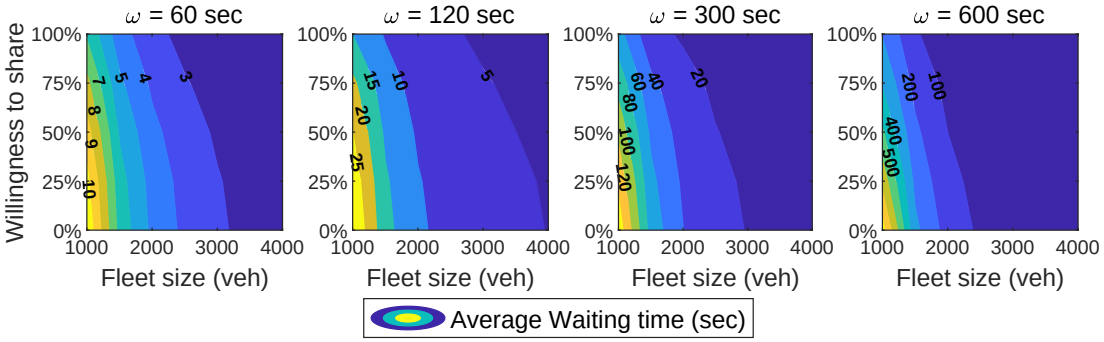


Figure 3.4: Average waiting time (including a penalty for abandonment) as a function of fleet size, willingness to share and waiting time tolerance (ω).

It is interesting how the sensitivity analysis shows the occurrence of shared rides. Figure 3.5 confirms that passengers' willingness to share is the most relevant parameter when computing the number of shared rides out of all provided ride-sourcing rides. Waiting time tolerances become relevant only when too small, severely reducing the number of shared rides. As one could expect from the assumption of matching incoming requests to the nearest driver and not prioritizing shared rides, larger fleets of active drivers decrease the number of shared rides.

A close look at the dynamics of specific instances reveals how some parameters change drivers' activities in the experiment. Figure 3.6 shows that regions 1 and 2 remain nearly one hour without idle drivers, losing incoming ride-hailing requests. Higher passengers' willingness

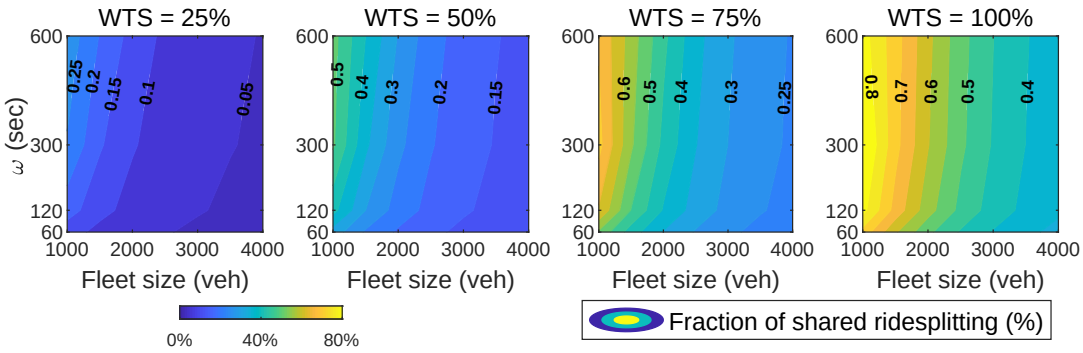


Figure 3.5: Fraction of shared rides from all rides as a function of fleet size, waiting time tolerance (ω) and willingness to share (WTS).

to share was unable to avoid such a situation but still managed to serve more passengers, reducing the number of ride-sourcing travelers switching to private vehicles in all regions. At the peak, the difference was around 1000 fewer private vehicles only in region 2, comparing scenarios of 0% and 100% willingness to share.

3.4 Comparison with a detailed event-based simulator

The road network for the Futian and Luohu districts of Shenzhen, China, forms the background for the study. The considered network consists of 1'858 intersections connected by 2'013 road segments. In total, the Origin-Destination demand data contained around 200'000 requests collected from taxi operations using GPS coordinates (Ji et al., 2014). The experiment used a simulator based on Beojone and Geroliminis, 2021b (Chapter 2), which had historical data translated into Table 3.3 to use in the evaluated forecasts. It also includes the MFD data in Figure 3.7 and a constant coefficient of variation $\sigma/L = 0.57$.

Table 3.3: Two-region model parameters.

OD-pair	Demand ratio	Transfer ratios (θ_{okd})		Passage ratio (β^k_{od})		$L^{\text{drop}}_{K_{od}}$ (km)		
		$k=1$	$k=2$	$k=1$	$k=2$	RH/PV	$S1$	$S2$
1-1	0.390	n/d	n/d	1	0	2.801	2.773	2.746
1-2	0.116	0	1	1	1	3.324	3.247	3.069
2-1	0.111	1	0	1	1	3.854	3.733	3.743
2-2	0.383	n/d	n/d	0	1	3.315	3.280	3.257

3.4.1 Simulation/Plant description

The simulation/plant consists of an event-based spatial traffic simulation based in Beojone and Geroliminis, 2021b. It tracks every new trip based on its geographical origin, destination, and traveled distance in an urban network designed as a graph of roads and intersections.

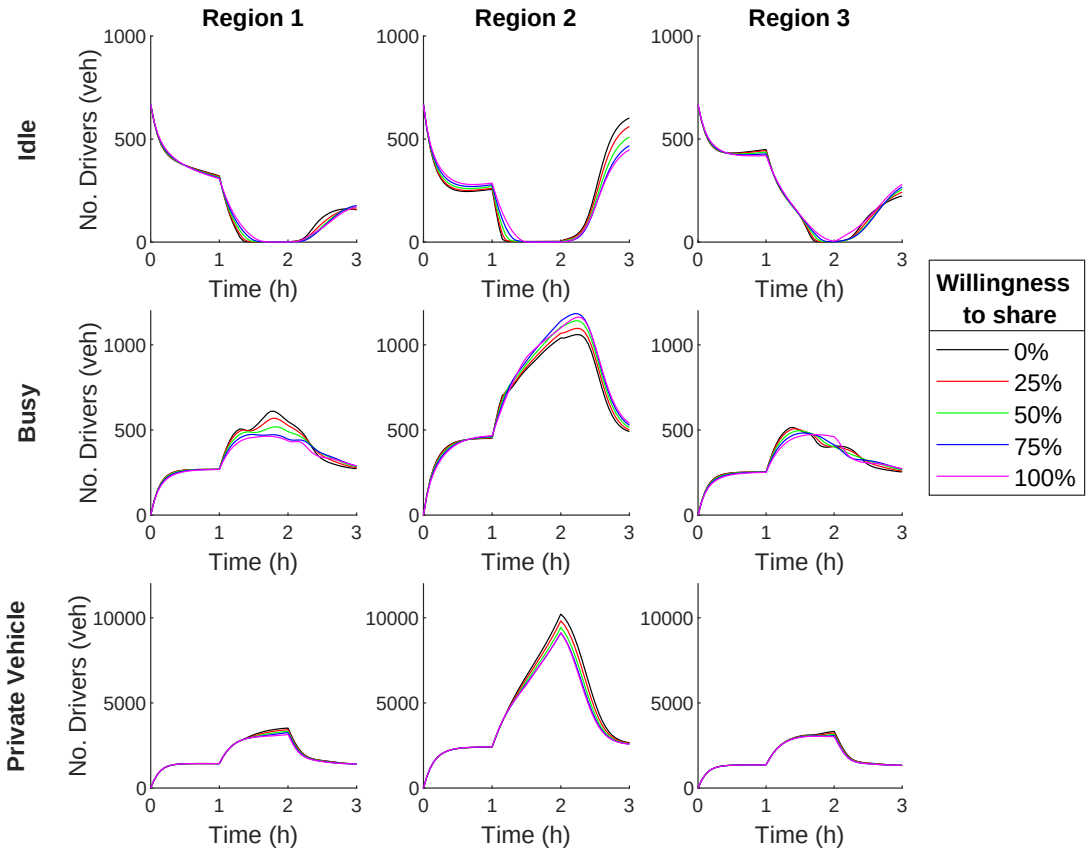


Figure 3.6: Summarized number of idle and busy drivers and private vehicles for the cases with 2000 ride-sourcing drivers and 10 minutes of waiting time tolerance.

Differently from classical trip-based models, vehicles have their microscopic geographical positioning tracked to evaluate detailed passenger-driver matching constraints for ride-hailing and ridesplitting. A Speed-MFD estimates time-varying speeds shared among all links of a region. The previous eliminates the expensive traffic assignment process, and vehicles may travel through the shortest path. To have accurate positions and passenger-driver matching evaluations, each entity in the simulation has a tuple of information characterizing them.

The arrival of a passenger marks the start of a ride-sourcing request. Waiting time and detour tolerances are set for all passengers. The matching process for ride-hailing requires an idle driver close enough to the arriving passenger to comply with the waiting time tolerance. For ridesplitting, besides idle drivers, those assigned to another ridesplitting passenger are potential assignments for arriving passengers. However, in these cases, the evaluation must also check whether the detour will be acceptable for both passengers. If a ride request is feasible, it is accepted and a ride-sourcing driver is assigned to pick-up and deliver the respective passenger.

Ride-sourcing drivers are responsible for picking up and delivering passengers in the modeled

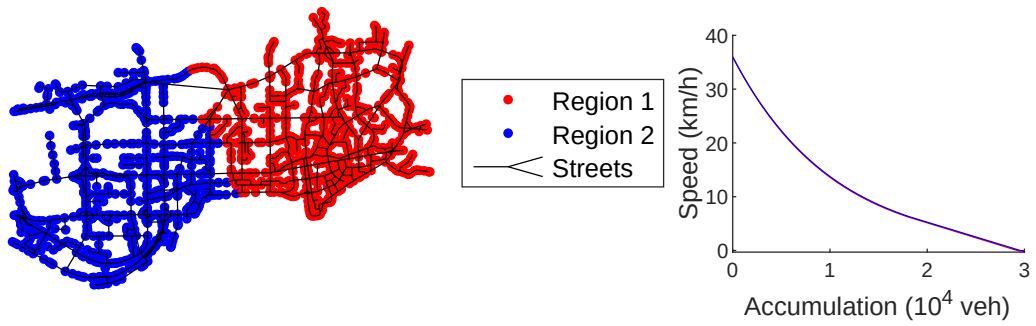


Figure 3.7: Shenzhen central business district separated in two regions and their respective Speed-MFDs used in the accuracy analysis.

road network. Tracking their positions and activities allows the simulation to check their availability for matching with arriving passengers and dispatching them accordingly. The simulation follows TNCs' common practice of assigning passengers to the closest available driver on an FCFS basis. Assignments determine the sequence of visited intersections for pick-ups and drop-offs. In the case of ridesplitting requests, they are ordered to minimize the total traveled distance as long as the detour tolerance is fulfilled for all involved passengers.

The majority of entities affecting traffic is the background traffic. A simpler tuple represents private drivers' situation and position. Once the driver reaches the destination, the vehicle leaves the system (by parking outside the road space, for instance). These entities do not interact with the ride-sourcing service, except for the lost ride-sourcing requests using private vehicles to fulfill their trip demand and traveling speeds.

Differently from Beojone and Geroliminis, 2021b, we separated the studied area into a set of regions. Moreover, both ride-sourcing drivers and private vehicles have additional properties to track their interregional path (including each intraregional trip length and sequence of regions in a trip). The shortest paths (distance and sequence of intersections) are defined using a Floyd-Warshall algorithm.

3.4.2 Error evaluation

In the error evaluation component, the simulation provides reference values, and the dynamic model provides the forecasts for comparison. However, depending on the application, the model must provide predictions for different time horizons. An MPC controller used for real-time fleet management needs several short-term predictions, and its efficiency relies on the quality of those (Sirmatel & Geroliminis, 2018a; Sirmatel & Geroliminis, 2021).

Therefore, we mimic an MPC controller using a rolling time horizon framework to evaluate the model. Every Δt time units, the simulation halts and describes the system, including information about ride-sourcing and private vehicle numbers and their respective remaining

distances. Then, from halt time t_i, the model forecasts the system's evolution for the next T steps of δt time units. Note that halting times t_i are Δt units apart from each other and $T\delta t \geq \Delta t$.

Firstly, the evaluation computes the error related to estimates of state K_{od} for T forecast steps starting at t_i, called $\varepsilon^K_{od}(t_i, T\delta t)$. Equation [3.26] illustrates the absolute error of the prediction for state K_{od}, measured as the "number of vehicles" (veh).

$$\varepsilon^K_{od}(t_i, T\delta t) = \sum_{l=1}^{T} \left| \hat{n}^K_{od}(t_i, l\delta t) - n^K_{od}(t_i + l\delta t) \right| \tag{3.26}$$

where $n^K_{od}(t)$ stands for the actual number of vehicles in state K_{od}; and $\hat{n}^K_{od}(t_i, t_i + \delta t)$ is the predicted value of $n^K_{od}(t_i + \delta t)$ when starting the prediction at t_i.

The relative error $\varepsilon(t_i, T\delta t)$ aggregates all errors for a given halting time t_i and prediction horizon $T\delta t$. The estimated error of the model is called $\varepsilon(T\delta t)$. It summarizes the errors for all halting times t_i depending on $T\delta t$. Equations [3.27] and [3.28] depict both of these dimensionless errors.

$$\varepsilon(t_i, T\delta t) = \frac{\sum_{K\in\mathbb{S}} \sum_{o,d} \varepsilon^K_{od}(t_i, T\delta t)}{\sum_{K\in\mathbb{S}} \sum_{o,d} \sum_{l=1}^{T} n^K_{od}(t_i + l\delta t)}, \qquad \mathbb{S} = \{I, RH, S1, S2, PV\}, \text{ and } \{o,d\} \in \mathscr{R}^2 \tag{3.27}$$

$$\varepsilon(T\delta t) = \sum_{t_i} \varepsilon(t_i, T\delta t), \qquad t_i = 0, \Delta t, 2\Delta t, ..., t_f, \text{ and } T = 1,2,3,4,5 \tag{3.28}$$

In this experiment, the simulation halts every $\Delta t = 3$ minutes (0.05h). Then the model predicts from 1 to 5 steps (T) of 6 minutes ahead of time, completing up to 30 minutes of forecasts. The settings for this experiment are illustrated by $\Delta t = 0.05$h, $t_i = 0, \Delta t, 2\Delta t, ...,$3h, $\delta t = 0.1$h, and $T = 1,2,...5$. We refer to this experiment as the "short forecast."

Other applications (e.g. pricing) might require longer predictions. Thus, it is necessary to understand its limitations and ability to describe system dynamics for different time horizons.

A second experiment consisted of a single model run for the entire evaluation period, receiving information from the simulation only at the beginning. For the remaining time, there is no information exchange between the model and the plant. In summary, the settings for this experiment has a single $t_i = 0$, $\Delta t = 3$h, and a $T\delta t = 3$h. We refer to this experiment as the "long forecast."

3.4.3 Benchmark models

To better emphasize the importance of having a more complex model for ride-sourcing dynamics compared to an accumulation-based MFD model, we utilized a benchmark model developed to model cruising for parking with MFD dynamics (Geroliminis, 2015). It was one of the first efforts to integrate dynamic trip lengths and a state representation that decomposes the trip of a vehicle to various components, as required given the features of ride-sourcing trips, but in an accumulation-based model. Simply speaking, the total production of the vehicles splits among the different states in a way analogous to the accumulations, and trip endings are estimated by dividing the specific production by the average trip length of the state (that can be time-dependent). Moreover, the literature contemplates dynamic models for taxi and ride-hailing services, which form a relevant benchmark for the proposed model. However, the presented model is distinctive for including ridesplitting operations deliberately. Therefore, we aggregated all ride-sourcing activities into a single busy state for the benchmark model. Ramezani and Nourinejad, 2018 also had a similar activity description for private and taxi vehicles separated into dispatched and occupied ones. In this approach, except for idle ride-sourcing vehicles, trip lengths are assumed constant and were estimated using the plant data. Trip lengths considered the entire distance a ride-sourcing vehicle traveled from its assignment to a passenger until the last passenger drop (becoming idle again), independently of the service option. We refer to this benchmark model as the "Acc.-based model" in the figures.

One could acknowledge that ride-sourcing operators are indifferent to traffic conditions when evaluating their service dynamics. For this reason, we wanted to evaluate the impact of tracking traffic conditions during predictions. To this end, we adapted this benchmark model to assume a constant free-flow speed $v_o(n_o(t) = 0)$ at the prediction horizon. In this case, we refer to it as the "Benchmark No-traffic". The benchmark's purpose is to highlight that even if TNCs might not be interested in the congestion their operations create, they should account for it if they are interested in managing their quality of service with real-time strategies (e.g., repositioning or surge pricing).

3.4.4 Model evaluation

The initial evaluation of the proposed model (prediction quality and stability, without measuring errors) consists of the experiments called "long forecast" and "short forecast," described in Section 3.4.2.

To illustrate the experiment mimicking an MPC controller ("short forecast"), Figure 3.8 shows three consecutive steps computed in a rolling time horizon prediction for idle ride-sourcing drivers in Region 2. The model predicts future system conditions every 3 minutes (0.05 hours). The model only considers the first 10 time steps (completing 30 minutes of forecasts) to reasonably use computational resources. The feedback loop from the plant to the prediction model estimates system states, including ride-sourcing and private vehicle information.

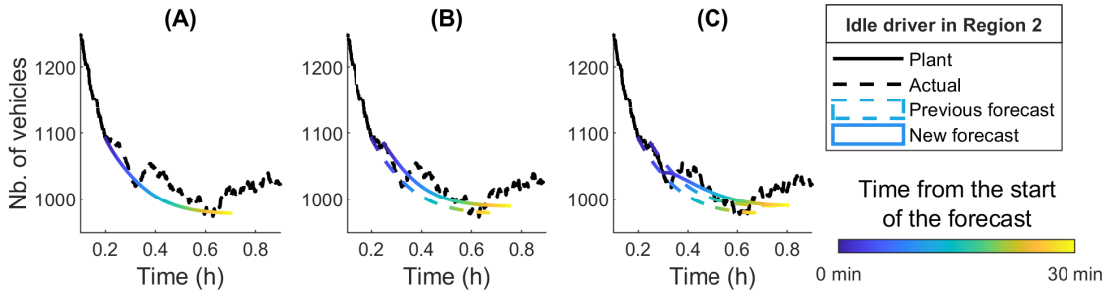


Figure 3.8: Rolling time horizon prediction instances for idle ride-sourcing vehicles in Region 2. Forecasts starting 12 (A), 15 (B) and 18 (C) minutes from the experiment beginning.

To assess whether the model can capture the state dynamic evolution, we separated a single round of the plant with the respective forecasts for the number of vehicles. Figure 3.9 depicts these measurements for short and long forecasts. Note that, for visualization purposes, we aggregated the number of vehicles according to their current region. We also aggregated all ride-sourcing vehicles with at least one assigned passenger into a "Busy" classification. In general, estimated values followed plant values closely for most states. In some short forecasts, such as $S1_{12}$, PV_{11}, PV_{21}, and most $S2_{od}$, the model initially moves away from the plant data, but it returns to the values close to the "long forecasts" and the plant values. The deviations remained, at most, in the order of 10^1 for ride-sourcing and 10^2 for private vehicles. The previous highlights that examining short forecasts may provide a better test of robustness. The regional number of idle and busy drivers is crucial for ride-sourcing operations. Various strategies require those. Some examples are vacant vehicle relocation, surge pricing, integration of ride-sourcing in High-Occupancy-Vehicle or High-Occupancy-Toll lanes, and perimeter control.

One of the central concerns in modeling ridesplitting activities is ensuring the model can capture key service characteristics, which are state- and demand-dependent. Figure 3.10 depicts the ratio of drivers $n_{od}^{S2}(t)/n_{od}^{S1}(t)$ on ridesplitting activities in each region. As one would expect, the more passengers joining ridesplitting, the more passengers have shared rides. In the beginning, few drivers carry multiple riders since, most times, an idle vehicle is the closest one to arriving passengers. However, once the demand grows, it forms the pool of drivers with a single ridesplitting passenger allowing for more shared rides, where hundreds of drivers in both regions have two simultaneous passengers. For instance, in Region 2, the number of drivers carrying two passengers almost equals the number of drivers carrying a single passenger at 2h. It shows a seven-fold increase in a period of 75% larger demand, highlighting the responsiveness to market thickness.

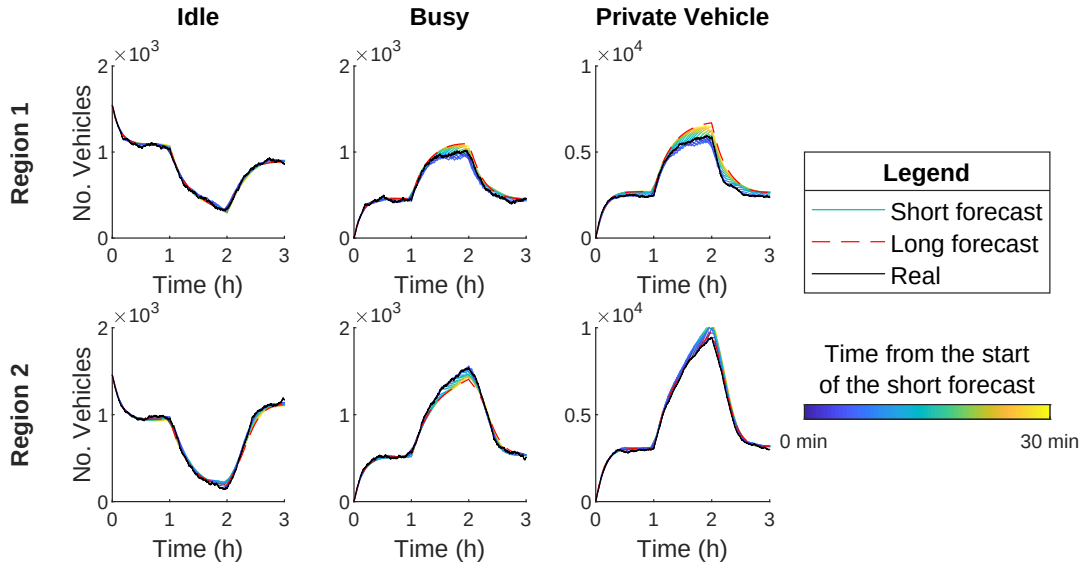


Figure 3.9: Results for long and short forecasts compared directly to the plant results. Model states are aggregated per current region and vehicle situation.

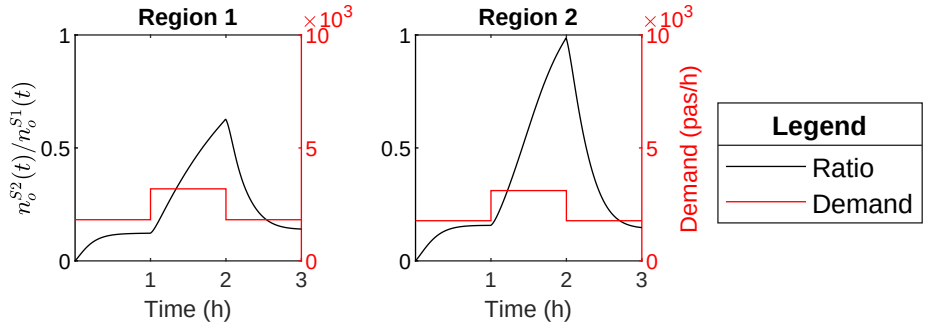


Figure 3.10: Ratio of shared rides compared to ridesplitting demand.

3.4.5 Benchmark comparison

Error measurements can provide a detailed analysis of the quality of the forecasts and how they deteriorate at later steps. In Figure 3.11, we compile the forecasts from 30 independent experiment runs. Firstly, in Figure 3.11 (Left), one can note that relative errors are naturally higher for longer predictions, as expected. Furthermore, errors often remained below 5%, even on forecasts of 12 minutes or more. Errors were higher than 6% only at the most crowded moments. Measurements on $\varepsilon(t_i, T\delta t)$ converge at 0 by the end of the experiment because there are no forecasts (nor plant data) after 3h. In Figure 3.11 (Right), we included total error measurements for all the benchmarks and the proposed model. Excluding 'Benchmark No-traffic', boxplots presented increasing variability in later forecast steps but remained small compared to the average. For instance, coefficients of variation ranged between 0.005 and

0.022 for the benchmark and the proposed models, respectively. Total error measurements, $\varepsilon(T\delta t)$, presented nearly linear increases for the number of steps. Errors of the 'Acc.-based MFD model' were nearly double those of the proposed model, while the 'Benchmark No-traffic' model marked around 5 times higher errors (plotted above the other models).

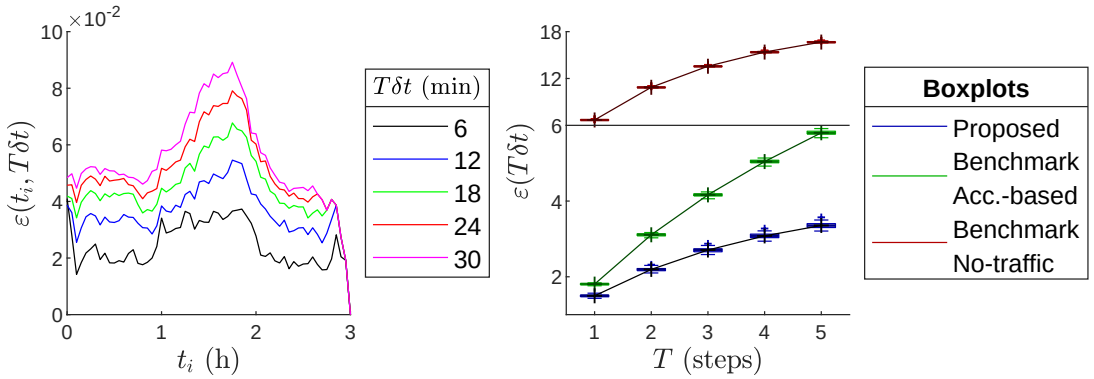


Figure 3.11: Summary of error measurements. (Left) Subtotal forecast errors for different $T\delta t$; and (Right) Boxplots of total errors according to the $T\delta t$.

We acknowledge that we computed error measurements in Figure 3.11 for all model states, not only the aggregated ones in Figure 3.9. Hence, errors for the proposed model accounted for all 18 states, while the 'Acc.-based MFD model' and the 'Benchmark No-traffic' accounted for only 10 states. Total and subtotal errors of the proposed model were inferior to those of all benchmarks. It confirms the proposed model as a better approximation to the traffic system.

To evaluate if the errors are distributed differently in each model, we separated them for all vehicles according to their situation and current region in Figure 3.12. To have a fair comparison, we aggregated the vehicles in groups before computing the errors ($\varepsilon^K_{od}(t_i, T\delta t)$) in all models. As shown in Figure 3.12, private vehicles represent most of the errors for all models. In the 'Acc.-based MFD model,' the errors have a similar share distribution as the demand, where 85% of it refers to private vehicles. The proposed model reduced the errors for these vehicles by 60%, even with the same modeled states. On the other hand, the proposed model has 8 states more for ride-sourcing vehicles, and errors reduce by between 28% and 42%. The total error dropped to less than half, indicating that the computation of trip completion and transfer flows is responsible for such results. At the same time, it provided more detailed information on ridesplitting operations. As the last evaluation on the dynamics of 'Benchmark No-traffic,' we checked whether the increased errors from Figure 3.11 concentrated in private vehicles. However, errors had a similar distribution as the other tested models, highlighting the importance of traffic dynamics in evaluating ride-sourcing operations (even if the operator is not interested in it).

As a note, we ran tests separating pick-up and delivery activities to evaluate possible shortcomings of the aggregation process. However, even with the additional description, it had slightly higher $\varepsilon(T\delta t)$ and $\varepsilon^K_{od}(t_i, 30\text{min})$ than the proposed model. Results were worse than

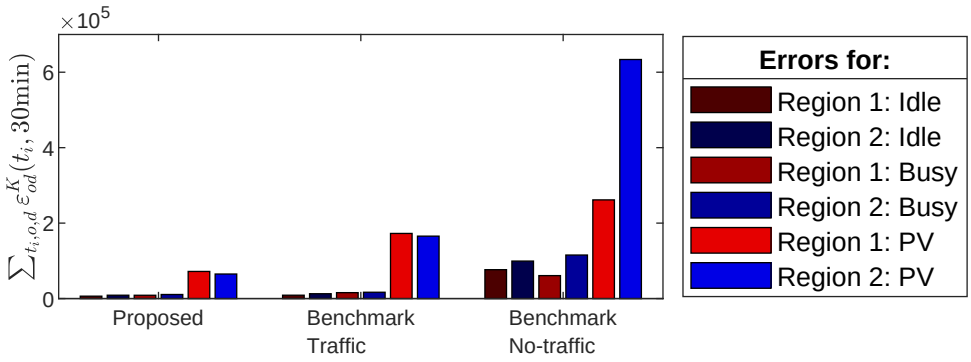


Figure 3.12: Summary of total errors in forecast of 30 minutes (5 steps). Total errors per vehicle situation and current region.

the proposed model because it is more susceptible to measurement noises, particularly those in pick-up activities. After all, they have very short average trip lengths.

3.5 Summary

In this chapter, we sought to develop a dynamic aggregated traffic network model capable of representing ride-sourcing services and background traffic in a macroscopic multi-region urban network.We combined the Macroscopic Fundamental Diagram (MFD) with detailed state-space and transition descriptions of background traffic and ride-sourcing vehicles in their activities to formulate mass conservation equations. Accumulation-based MFD models might experience additional errors due to the variation profile of trip lengths, e.g., when vehicles cruise for passengers. We integrate the so-called M-model that utilizes the total remaining distance to capture dynamics of regional and inter-regional flows and accumulations for different vehicle (private or ride-sourcing) states. This aggregated model is capable to reproduce the dynamics of complex systems without using resource-expensive simulations. We also show that the model can accurately forecast the vehicles' conditions in near-future predictions (e.g., 30 minutes ahead). Later, a comparison with benchmark models shows lower errors in the proposed model in all states. Finally, we evaluate the model's robustness to noises in its inputs, and forecast errors remain below 15% even where inputs are 20% off the actual values for ride-sourcing vehicles. The development of this model prepares the path for developing real-time feedback-based management policies such as priority-based perimeter control or repositioning strategies for idle ride-sourcing vehicles and developing regulations over ride-sourcing in congested areas.

4 Guiding the relocation of ride-sourcing drivers with revenue forecasting

This chapter is based on the following papers:

- Beojone, C. V., & Geroliminis, N. (2023c). Relocation incentives for ride-sourcing drivers with path-oriented revenue forecasting based on a markov chain model. *Transportation Research Part C: Emerging Technologies, (Under Review)*
- Beojone, C. V., & Geroliminis, N. (2023b). Providing a revenue-forecasting information scheme as an incentive to relocate compliant ride-sourcing drivers. *102nd Transportation Research Board (TRB) Annual Meeting*, 1–15

4.1 Introduction

It is in the best interest of the Transportation Network Company (TNC) responsible for the service operation to balance demand and supply to maintain satisfactory service quality. However, the fleets of this service are formed by human drivers that offer rides to make profits and are free to perform a series of decisions, which include defining where they will look for new assignments and when to offer rides. Therefore, the TNC cannot deploy extra vehicles whenever there is a shortage, and it involves convincing power to relocate the available pool of drivers.

Herein, in this chapter, we evaluate the potential repositioning response of drivers when provided with an estimate of their earnings. Drivers are not forced to perform specific actions and may not comply with the provided guidance. They base their decisions on their earning expectations for remaining idle in the current region or repositioning to a neighboring one. The operator uses a mixed discrete-continuous time Markov chain (MDCTMC) to estimate individual earnings for a given decision in the short term. A microscopic process identifies the positions and associated paths that provide the highest chances of matching. In a simulated

study based in Shenzhen, China, we compare the performance of guided and unguided drivers.

Following the motivation and the challenges raised regarding improving the ride-sourcing service through repositioning, which is given in Section 1.2.3 of Chapter 1, the remainder of the chapter is structured as follows. Sections 4.2 and 4.3 describe the proposed framework to identify the best repositioning decisions and to persuade drivers by predicting their activities and earnings. They first depict the optimal microscopic decisions (intersection- and path-wise), then describe the proposed MDCTMC and the revenue estimation process. Section 4.4 shows the computational results in the simulated study and the comparisons to benchmark models. Finally, Section 4.5 summarizes the main findings.

4.2 Identifying the best repositioning decision

While drivers decide about repositioning by themselves, intending to maximize their profit, they have incomplete information about the system conditions. They also have limited rationality for a complex problem with thousands of other vehicles, preventing them from finding the best repositioning decisions. Therefore, it is more likely that an operator who agglomerates information about all drivers and knows better the spatiotemporal characteristics of demand will have the means to identify repositioning needs and provide guidance.

It is worth mentioning that the service provider can use the mobile application to supply drivers with such information since it has access to historical and real-time data about the demand and other operational data, such as the spatial distribution of drivers over the city. The idea would be to utilize this data and identify the best repositioning decisions and present the driver with the expected best ones when the driver finishes a trip with a passenger and becomes available for relocation. Hence, the driver receives detailed information about the place and path they should drive as soon as they complete their last request (becoming idle). We do not follow a system-optimum-oriented approach, as this might create more tangible issues with driver compliance, and its implementation can be more challenging.

Since drivers offer rides in ride-sourcing services for a profit, one reasonable expectation of drivers' motivation is profit maximization. Assuming that a driver maximizes his/her revenues by maximizing the number of served requests, a driver's repositioning decisions should increase his/her chances of getting new assignments.[I] To encapsulate this idea, one must consider the passenger-driver matching process, where we assume the urban area is partitioned into different regions. Without loss of generality, we assume an instantaneous matching policy, where passengers are assigned to the closes available driver in their region on a first-come-first-served basis. Additionally, the platform temporarily limits the assignments to repositioning drivers to ensure the execution of a repositioning decision. For instance, a driver cannot be assigned to a passenger unless the latter has the same destination region as

[I] In reality, some requests might be more profitable than others, and a driver could maximize his/her revenues while serving fewer requests. That does not change the idea that, in general, serving more requests will increase one's revenues.

the driver. Finally, once entering the final repositioning destination region, these restrictions are lifted, even before reaching the exact oriented location.

In summary, the chances of matching a driver depend on his/her and other drivers' positions. Since one can expect drivers to concentrate in areas of higher demand (at least until reaching an overcrowded level), a driver repositioning through poorly covered areas can have an additional chance of receiving an assignment. That is, poorly covered areas can be a better repositioning option than one with higher demand.

Given the fast-changing conditions in ride-sourcing services, the vehicle density in one area can vary significantly in a short period. Therefore, repositioning decisions must account only for the near future, where a significant portion of the prediction is spent moving from the current position i to a potential destination j. Then, the chances of getting assigned vary significantly during the prediction horizon and can pose a challenge in identifying the best repositioning destination (and its associated path).

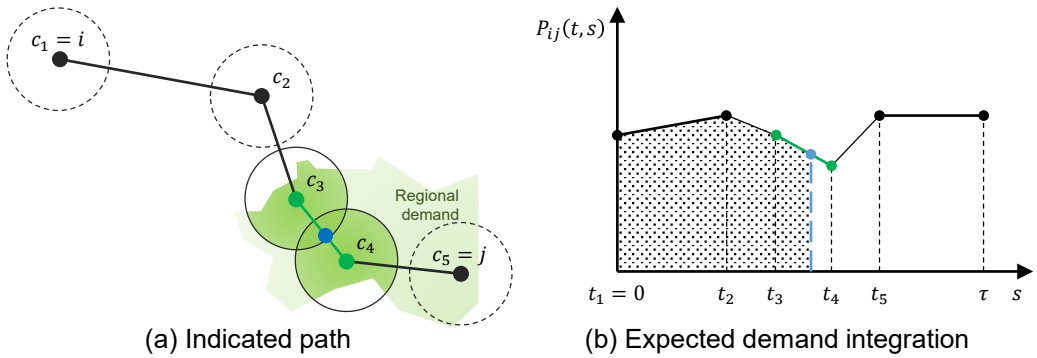
4.2.1 Path orientation solution

Assume that all trips originate and finish at intersections of the road network. We can identify the destinations with the highest number of expected assigned requests from a finite (and discrete) set of destinations. Since the number of intersections in one region is relatively small, the decision set is small, allowing us to perform ranking and selection of the best destination in each region. Note that, without loss of generality, we arbitrarily suppressed the different service options (ride-hailing and ridesplitting) and the regional description from the notation in this section.[II]

Let us consider a path for a driver as a sequence of nodes between intersections i and j, as shown in Figure 4.1a, where i is the closest intersection to a driver's current position and j is another intersection evaluated as a potential destination in the street network. Assume that an expected number of assigned requests $P_{ij}(t)$ with this path starting at time t is the result of the integration of the instantaneous assignment rate $P_{ij}(t,s)$ given by Equation [4.1], where $0 \leq s \leq \tau$ is the time elapsed from the beginning of the forecast, and τ the prediction horizon. To calculate the instantaneous $P_{ij}(t,s)$, we should consider other idle vehicles and their coverage, such that the assignment rate accounts for the competition among drivers. Note that it does not mean that the driver will be assigned all $P_{ij}(t)$ requests; it is a reference to identify the path in which the driver is most likely to find a new assignment. The next Section 4.2.2 develops an approximation for $P_{ij}(t,s)$.

$$P_{ij}(t) = \int_0^{\tau} P_{ij}(t,s)\,\mathrm{d}s \tag{4.1}$$

[II]To compute the number of expected requests from both services, one must repeat (or do it in parallel) the steps described in the following subsection.



(a) Indicated path (b) Expected demand integration

Figure 4.1: Illustration of the elements in the computation $P_{ij}(t)$. (a) indicated path, where circles around nodes c_1 to c_5 represent the demand covered. (b) Depiction of the demand integration associated with the indicated path.

Hence, with the computation of $P_{ij}(t)$ for each potential destination, the solution of the proposed path optimization can be summarized into Equation [4.2]. The idea is to rank each $P_{ij}(t), \forall j$, and select the potential destination node j returning the highest position in the ranking, i.e., with the maximum expected number of assignments in the prediction horizon. In a multi-region setting, one could identify the optimized destination for each region by separating the evaluated intersections in each area (Equation [4.3]). In the next section 4.2.2 we detail how we use this elegant framework to simulate all possible solutions and rank them to find the best solution to Equations [4.2] and [4.3].

$$\Pi_i(t) = \max_j P_{ij}(t) \tag{4.2}$$

$$\Pi_i^r(t) = \max_{j \in r} P_{ij}(t), \qquad \text{where } r \in \mathscr{R} \tag{4.3}$$

where r is a region from the set of regions $\mathscr{R}$ in which the urban area is partitioned.

Note that this approach will not guide drivers to areas crowded with idle drivers or with a path that provides few matching opportunities. It balances the distance and the demand on a driver's path to maximize the probability of an assignment within the prediction horizon.

4.2.2 Assignment rate associated with the oriented path

Since the operator should provide guidance quickly to available drivers, this process should be fast to compute. Therefore, instead of using a detailed simulation to enumerate all vehicles' movements and possible demand arrivals, we develop an elegant framework that uses current information and general predictions of future demand (e.g., demand arrival rate in one area and the historical geographic demand distribution), assuming a spatially distributed Poisson

process (Yu et al., 2019). The framework can be summarized in the following steps:

- Step 1: Identifying all drivers' latest destinations (including non-idle ones);
- Step 2: Computing a matching radius for all drivers from their latest destinations;
- Step 3: Checking the number of drivers covering each demand node (from their latest destinations);
- Step 4: Uniformizing demand based on the coverage;
- Step 5: Summing the uniformed demand rate covered in each demand node;
- Step 6: Computing the assignment rate associated with the path to each potential destination;

Following the latest destination in Step 1, Step 2 computes the matching radius for all drivers idle or busy. To this end, we assume that vehicles will only cover demand from their latest destination (Step 1), i.e., where they completed their last request or where they are heading in their current activity.[III] Additionally, to this assumption, only requests that can be assigned within the prediction horizon τ are considered. Thus, the coverage radius at their destination can decrease if the driver is too far from it (possibly reducing the radius to zero). Note that this avoids counting unfeasible demand coverage for drivers that will not reach that area in time to compete for requests with the evaluated driver. Equation [4.4] illustrates the computation of the matching radius following this assumption.

$$R_{id,\mathrm{d}_{id}}(t) = \min\left(\omega, \tau - T_{id,\mathrm{d}_{id}}(t)\right) \cdot v_{\mathrm{d}_{id}}(t) \tag{4.4}$$

where ω is a passenger waiting time tolerance and $T_{id,\mathrm{d}_{id}}(t)$ is the travel time for vehicle id to reach its current destination d_{id} under current conditions (at time t). $v_{\mathrm{d}_{id}}(t)$ is the instantaneous travelling speed around d_{id}.

For Steps 3 and 4, we assume that all vehicles covering a demand node are equally likely to be assigned to a request from that node.[IV] This simplification decreases the computational effort and encapsulates the uncertainty related to other drivers' cruising behavior. Therefore, Equation [4.5] illustrates the counting process for drivers covering requests from a node c. Next, Step 4 uses the information from Step 3 to uniformize the demand rate originating from each node (Equation [4.6]).

[III] This assumption can become quite constraining, but keeping track of the positions of all vehicles throughout τ would be too costly to integrate into such a framework.

[IV] A more detailed alternative to this assumption would be to compute a Voronoi diagram that varies in time, accounting for the drivers' movements and positions.

$$n_c(t) = \left|\left\{id : \text{dist}(\text{d}_{id}, c) \leq R_{id,\text{d}_{id}}(t)\right\}\right| \tag{4.5}$$

$$p'_c(t) = \frac{p_c(t)}{n_c(t)+1} \tag{4.6}$$

where the $|\{\cdot\}|$ indicates the cardinality of a set; $\text{dist}(\cdot,\cdot)$ is the distance between two points; and $p_c(t)$ and $p'_c(t)$ are estimates of demand arrival rate before and after the uniformization, respectively.

Figure 4.2 compares the repositioning decisions in one specific region when this process is not utilized to highlight the need for uniformization of demand rates in Step 4. Note that uniformization represents how densely covered the demand in some areas is. For illustrational purposes, the coverage radius is represented as a circle (indicating Euclidean distances), but the simulation used network distances instead. The figure highlights a difference in the suggested destination (the solution method is expanded in the next Section 4.2.1). Non-uniformized demand leads drivers to the area of highest demand because it overestimates the expected number of assignments in busier areas. On the other hand, uniformized demand leads to a closer destination at an area with a lower demand rate but poorly covered.

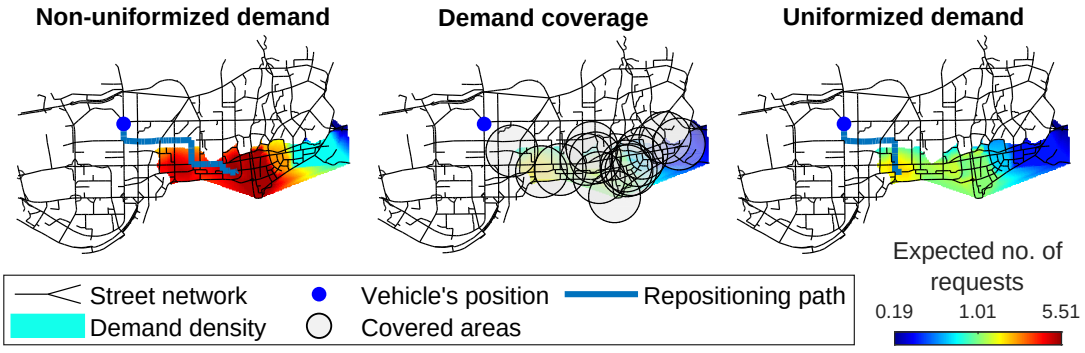


Figure 4.2: Comparison of potential repositioning decisions with and without Step 4 (covered demand uniformization).

In Step 5, we sum the total uniformized demand rate covered in each node. As illustrated in Equation [4.7], the coverage depends on a waiting time tolerance ω and the traveling speeds nearby the observed node c. Figure 4.3 depicts the elements in Step 5, where the covered demand is that within a reachable area and fulfilling any other passenger-driver matching constraints. Then, $P_c(t)$ becomes the sum of the uniformized demand rates covered by a vehicle at node c.

$$P_c(t) = \sum_{q:\text{dist}(q,c)\leq\omega v_c(t)} p'_q(t) \tag{4.7}$$

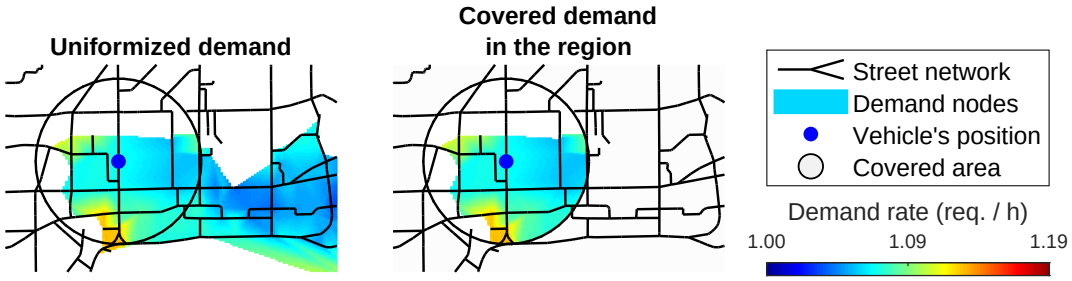


Figure 4.3: Illustration of the elements in the computation of $P_c(t)$. Highlighting only the demand in the current region.

Finally, Step 6 compiles the information from previous steps to compute the assignment rate associated with a repositioning destination $P_{ij}(t,s)$. The first part constructs the path C_{ij} between current position i and potential destination j and the associated travel times T_{ij} required to reach each node in that path. The constructed path C_{ij} is formed by a sequence of nodes in the path connecting i and j. We use a shortest path algorithm (e.g. Dijkstra's algorithm) to compute the sequence $C_{ij} = \{c_1 = i, c_2, \ldots, c_n = j\}$. In the same fashion, the associated travel times sequence T_{ij} is formed in ascending order. Therefore, for a sequence $T_{ij} = \{t_1 = 0, t_2, \ldots, t_n\}$, we have $0 < t_1 < t_2 < \cdots < t_n$. Based on the shortest path, we can iteratively compute t_x as shown in Equation [4.8].

$$t_x = t_{x-1} + \frac{\text{dist}(c_{x_1}, c_x)}{v_{c_x}} \tag{4.8}$$

Then, the dynamic assignment rate along this path $P_{ij}(t,s)$ is defined in Equation [4.9] as a linear interpolation between demand rates, such that $t_x < s \leq t_{x+1}$, and $t_x, t_{x+1} \in T_{ij}$.[V]

$$P_{ij}(t,s) = P_{c_x}(t) + \frac{P_{c_{x+1}}(t) - P_{c_x}(t)}{t_{x+1} - t_x}(s - t_x) \tag{4.9}$$

Nevertheless, we must highlight that the sequences C_{ij} and T_{ij} are unrelated to the prediction horizon τ so far. Thus, the driver can reach the designated position before the end of the prediction horizon ($t_n < \tau$), turning the interpolation from Equation [4.9] problematic. In that case, before computing $P_{ij}(t)$, we assume that the driver will remain nearby his destination j, thus, covering the same demand from node j, $P_j(t)$, until the end of the prediction horizon. In practical terms, before computing $P_{ij}(t)$, we append j again to the end of the sequence C_{ij} at the same time that we append τ at the end of T_{ij}.

[V] Given the non-smooth nature of $P_{ij}(t,s)$, Equation [4.1] is integrated numerically in the prediction horizon τ, using a typical trapezoidal method. Note that the sequence T_{ij} provides a list of steps, and $P_c(t), \forall c \in C_{ij}$ provides the required heights in the numerical integration process.

4.3 Forecasting a driver's activities and revenues

The operator will present drivers with the expected earnings associated with repositioning options, offering drivers the opportunity to decide. Driver compliance is not guaranteed; nevertheless, the guidance should persuade drivers to perform decisions that improve service quality while improving their earnings. We assume that the drivers might comply with a probability that is estimated by a logit model. We also compare with a full compliance scenario.

To forecast the earnings of potential repositioning decisions (as those raised from previous Section 4.2), the service operator must identify the drivers' activities and actions in the near future. We describe drivers' activities based on their ongoing assignments, either ride-hailing or ridesplitting, as detailed below.

- Vacant (I): a vacant driver available for passengers.
- Ride-hailing (RH): a busy driver assigned to a ride-hailing passenger.
- Single ridesplitting ($S1$): a driver assigned to a single ridesplitting passenger.
- Shared ridesplitting ($S2$): a driver assigned to two ridesplitting passengers.

We predict drivers' actions by identifying the time spent at each activity given the demand, service conditions, and potential repositioning decisions. Besides activity duration, one must depict how a driver transitions among different activities and regions.

Consider a list of activities $\mathbb{A}$, such that $A \in \mathbb{A}$ indicates a driver's current activity. A set $\mathscr{R}$ with R heterogeneous regions, i.e., $\mathscr{R} = \{1, 2, ..., R\}$ illustrate the urban network area, while the pair $od \in \mathscr{R}^2$ depicts a driver's current and destination regions. Therefore, $A_{od} \in \mathscr{K}$ describes a driver current state in the set of all possible states. Setting the list of activities $\mathbb{A} = \{I, RH, S1, S2\}$ assumes that a driver can execute the following four activities completing the state-space with a size $|\mathscr{K}| = |\mathbb{A}| \cdot |\mathscr{R}|^2$.

The predictions assume a *first-come-first-served* (FCFS) assignment discipline where arriving passengers are assigned immediately to the closest available driver. Additionally, it assumes a maximum capacity of two simultaneous passengers in a shared ridesplitting ride, and a vacant driver can be either waiting for a new assignment in their current region (state I_{oo}) or repositioning to another region (state I_{od}, where $o \neq d$).

4.3.1 Markov Chain model

Assuming that the time spent in each activity is exponentially distributed, a Markov chain depicts most of a driver's movements. With detailed information on the urban area and the driver following the provided path (Section 4.2), the operator accurately estimates how long this driver needs to reach different regions during the repositioning. Therefore, the time spent

in repositioning activities is assumed to be deterministic.Consider a driver who is moving according to the path between i and the destination j^* (the optimal solution from Equation [4.2] or [4.3]), as defined in Section 4.2. Then, reorganizing the sequence of nodes C_{ij^*} to its regional level, we get $C^r_{ij^*} = \{r_1, r_2, \ldots, r_n\}$ representing the sequence of regions in the path ij^*, such that $i \in r_1$ and $j^* \in r_n$. In the same direction, T_{ij^*} can be reorganized to represent the time required to enter each region in $C^r_{ij^*}$, such that $T^r_{ij^*} = \{t^r_1, t^r_2, \ldots, t^r_n\}$.

As mentioned earlier, we assume an instantaneous matching policy where passengers are assigned to the closest available driver in their region on a *first-come-first-served* basis and the platform ensures the execution of a repositioning decision by limiting repositioning drivers to assignments matching his/her destination region. The matching restrictions related to the destination region are lifted after enough time to reach region r_n.

Based on the previous, we break down a driver's activity predictions into three phases: phase (a), representing the movements of a driver before reaching the destination region (subject to matching constraints while repositioning); phase (b), representing the driver reaching the boundaries on the repositioning path (becoming available to requests in the newly entered area while becoming unavailable in the previous one); and phase (c), representing the driver after completing the regional repositioning by first reaching the destination region r_n (free of any matching constraints related to repositioning).

Therefore, given the matching constraints and repositioning movements, Figure 4.4 illustrates how each one of these phases communicates to each other to predict a driver's actions, highlighting that each phase requires different dynamics to represent them. As illustrated in Figure 4.4, repositioning movements in phase (a) happen in the continuous time between regions in the repositioning path. Therefore, phase (a) uses continuous-time Markov chains (CTMC). Without the matching limitations and not constraining the driver's movements into a subset of regions, phase (c) also takes place in continuous time. Hence, we use another CTMC to represent phase (c). To represent phase (b), the instant in time when a driver changes regions, we use a discrete-time Markov chain (DTMC). In summary, the probability vector at the end of a phase becomes the starting condition of the next phase. In the next subsections, we detail the dynamics for each phase. Note that the construct that emerges from organizing these phases with different dynamics is named a mixed discrete-continuous time Markov chain (MDCTMC) model (Ingolfsson, 2005; Ingolfsson et al., 2007). We must highlight that such a structure enables individualized forecasting of activities and revenues since sequences $C^r_{ij^*}$ and $T^r_{ij^*}$ are based on the path provided to that specific driver, accounting for his/her position (Section 4.2). Finally, phases (a) and (b) are computed $|C^r_{ij^*}| - 1$ times before reaching region r_n and starting phase (c). In the case of a decision to stay in the current region, the amount of time required to fulfill this decision is negligible ($r_1 = r_n \therefore C^r_{ij^*} = \{r_1\}) \therefore |C^r_{ij^*}| - 1 = 0$). Therefore, this driver instantaneously passes from phases (a) and (b), moving directly to phase (c).

Previous studies indicate that the 'overhead' associated with the DTMC is small when com-

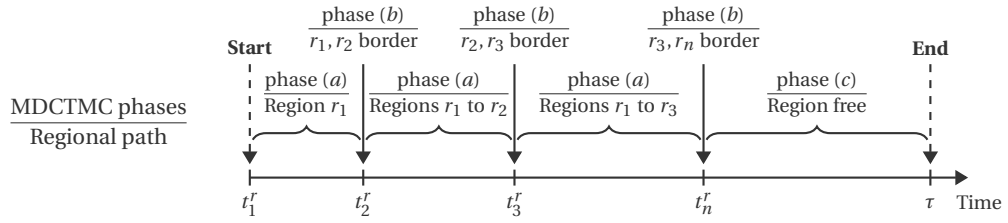


Figure 4.4: Illustration of the MDCTMC phases and its relation with the regional repositioning path in the forecasting timeline.

pared to the numerical solution of the differential equations in the other CTMCs of phases (*a*) and (*c*) (Ingolfsson et al., 2007).

Repositioning movements: phase (*a*)

Consider the time t, such that $t_l^r < t < t_{l+1}^r$ and $1 \leq l < n$. Therefore, if the driver received no assignments since the start of the repositioning activity, the driver would be in the l-th step in the regional repositioning path $C_{ij^*}^r$, r_l, but not yet in the destination region r_n. Recalling that the guidance provided to a driver is based on the shortest path, it is safe to assume that there is no case where the driver would be farther than region $r_l \in C_{ij^*}^r$. By extension, if the driver deviated from the shortest path due to any new assignments between the prediction start t_1^r and current time t, then this driver could be in any of the regions in the path between r_1 and r_l. Hence, we can limit the state space in the CTMC of phase (*a*) not further than the region r_l and its transitions include those from the starting region r_1 until the current region r_l, as depicted in Figure 4.5. Note that the matching constraints are applied, limiting the driver to assignments in his/her current region and those that the trip will end in the region r_n. Phase (*a*) ends at the moment the driver would reach the next region r_{l+1} at t_{l+1}^r, in the case of not being matched.

Equation [4.10] details the dynamics for all state probabilities $\pi_{wr_n}^K(t), \forall\, w \in \{r_1, r_2, \ldots, r_l\} \subset C_{ij^*}^r$, i.e., w is one of the regions in the regional path $C_{ij^*}^r$ preceding region r_l, included. Note that the dynamics include the probability of the driver being busy (with an assignment) not just in the region r_l but also in the previous ones because of the detours from the shortest path provided. Note that the structure of the MDCTMC ensures that $\pi_{wr_n}^I(t) = 0, \forall\, w \in \{r_1, r_2, \ldots, r_{l-1}\}$, therefore the transitions departing from these states are null (reason for not illustrating them in Figure 4.5) and, thus, $\pi_{wr_n}^I(t) > 0 \Leftrightarrow w = r_l$.[VI]

[VI] See the description of phase (*b*) for further details.

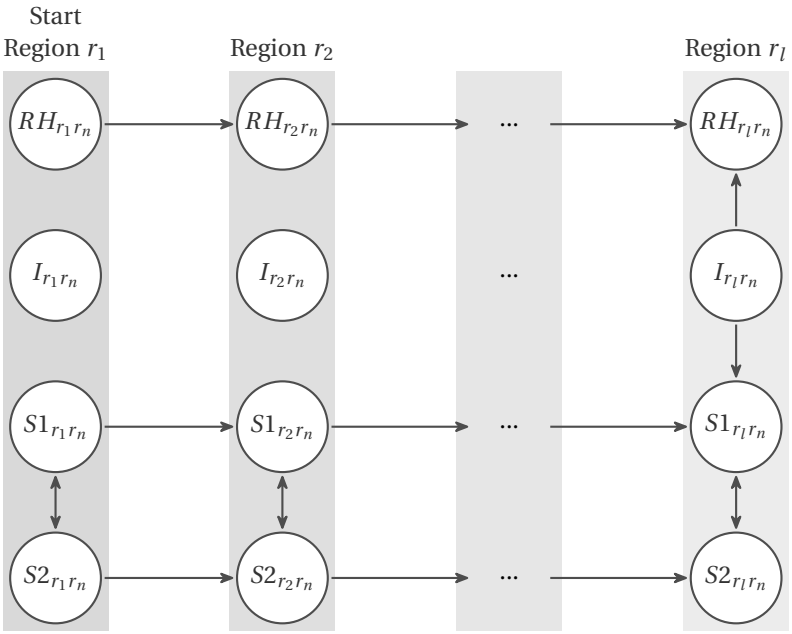


Figure 4.5: State-space transitions of the CTMC during the repositioning movement until region r_l in the repositioning path ($r_l \neq d$ and $l < n$).

$$\dot{\pi}^{I}_{wr_n}(t) = -\pi^{I}_{wr_n}(t) \sum_{s \in \mathscr{S}} \lambda^{s}_{wr_n}(t) \tag{4.10a}$$

$$\dot{\pi}^{RH}_{wr_n}(t) = -\pi^{RH}_{wr_n}(t)\mu^{RH}_{wr_n}(t) + \pi^{I}_{wr_n}(t)\lambda^{H}_{wr_n}(t) \tag{4.10b}$$

$$\dot{\pi}^{S1}_{wr_n}(t) = -\pi^{S1}_{wr_n}\left(\mu^{S1}_{wr_n}(t) + \sum_{h \in C^{r}_{ij^*}} \beta^{h}_{wr_n}\lambda^{S}_{wh}(t)\right) + \pi^{I}_{wr_n}(t)\lambda^{S1}_{wr_n}(t) + \pi^{S2}_{wr_n}(t)\vartheta_{wwr_n}(t)\mu^{S2}_{wr_n} \tag{4.10c}$$

$$\dot{\pi}^{S2}_{wr_n}(t) = -\pi^{S2}_{wr_n}(t)\mu^{S2}_{wr_n}(t) + \sum_{h \in \mathscr{R}} \hat{\lambda}^{S2}_{whr_n}(t) \tag{4.10d}$$

Where the base transitions can be described through the arrival process $\lambda^{s}_{wr_n}(t)$ for a service $s \in \mathscr{S} = \{H, S\}$ (ride-hailing and ridesplitting), and the service process $\mu^{K}_{wr_n}(t)$, in which a driver completes a ride or transfers to a neighboring region. Note that any $\mu^{K}_{wr_n}(t)$ and $\lambda^{s}_{wr_n}(t)$ is only defined for $w \in \{r_1, r_2, \ldots, r_l\}$, otherwise they have a value of 0. $\beta^{h}_{wr_n}$ represents the ratio of wr_n trips that will pass through region h. Therefore, $\beta^{h}_{wr_n}\lambda^{S}_{wh}(t)$ indicates the assignments of shared ridesplitting rides where the last passenger to board will be the first one to leave the vehicle, ensuring that the driver will end the trip in the region r_n. ϑ_{wwr_n} indicates the probability of the driver in $S2_{wr_n}$ having a passenger to deliver in w before proceeding to r_n with the other one.

Repositioning boundary: phase (b)

Phase (b) is active at the predicted time t_l^r the driver reaches the boundary between regions r_{l-1} and r_l. If the driver did not receive any assignments (leaving the state $I_{r_{l-1}r_n}$), then it means the driver is still repositioning and enters the following region r_l in the path $C_{ij^*}^r$ to continue with the repositioning in a new state $I_{r_l r_n}$. The previous describes the DTMC of phase (b), while Equation [4.11] depicts it supported by Equations [4.12] and [4.13] detailing the transition matrix. Note that the only change from the end of phase (a) occurs from state $I_{r_{l-1}r_n}$ to $I_{r_l r_n}$, whereas the remaining states keep unaltered probabilities. In other words, the driver could have reached the l-th step in the repositioning path, the neighboring region r_l, only if the driver remained in the repositioning activity $I_{r_{l-1}r_n}$ (traveling in the shortest path). Furthermore, since the travel time to reach the region r_l is deterministic, the transition occurs with certainty (probability of 1). If the driver started other activities, there must be at least a minimal detour from the shortest path to pick up the new assignment, therefore, it would not have changed regions at that precise moment (recalling the assumption of exponentially distributed times in the driver's activities).

$$\pi(t^+) = B(r_{l-1}, r_n, r_l)\pi(t^-) \tag{4.11}$$

$$B(r_{l-1}, r_n, r_l) = \left[b_{hk}^K(r_{l-1}, r_n, r_l) \right] \in \mathbb{B}^{|\mathscr{K}|\times|\mathscr{K}|} \tag{4.12}$$

$$b_{hk}^K(r_{l-1}, r_n, r_l) = \begin{cases} 1, & \text{for } K_{hk} = I_{r_l r_n}, RH_{r_{l-1}r_n}, S1_{r_{l-1}r_n}, S2_{r_{l-1}r_n}, S2_{r_{l-1}r_n} \\ 0, & \text{otherwise} \end{cases}, \tag{4.13}$$

Where t^+ and t^- refer to the instant right after and right before time $t = t_l^r$. $B(r_{l-1}, r_n, r_l)$ is the transition matrix representing the DTMC. $\mathscr{K}$ is the state space of the model, and $|\mathscr{K}|$ is its cardinality.

After repositioning: phase (c)

Recalling that once the time expected to reach the region intended in the repositioning decision t_n^r has elapsed, the driver becomes free of any limitations in the matching process imposed on repositioning vehicles. Therefore, Figure 4.6 illustrates the state space of a single region and the possible state transitions inside the depicted region and inter-regional movements. In the figure, regions 'k' and 'h' can be a set of regions of which the driver can pass immediately before and after region 'o,' respectively.

We describe the CTMC with the summarized Equation [4.14], where we estimate the state probability $\pi_{od}^K(t)$. Table 4.1 provides the entries for the summarized terms of the CTMC in the Equation [4.14]. Table 4.1 and Figure 4.6 do not show transitions related to repositioning states I_{od} because the structure of the MDCTMC ensures the completion of repositioning activities before phase (c). In other words, regions o, d, k and h in phase (c) are unrelated to

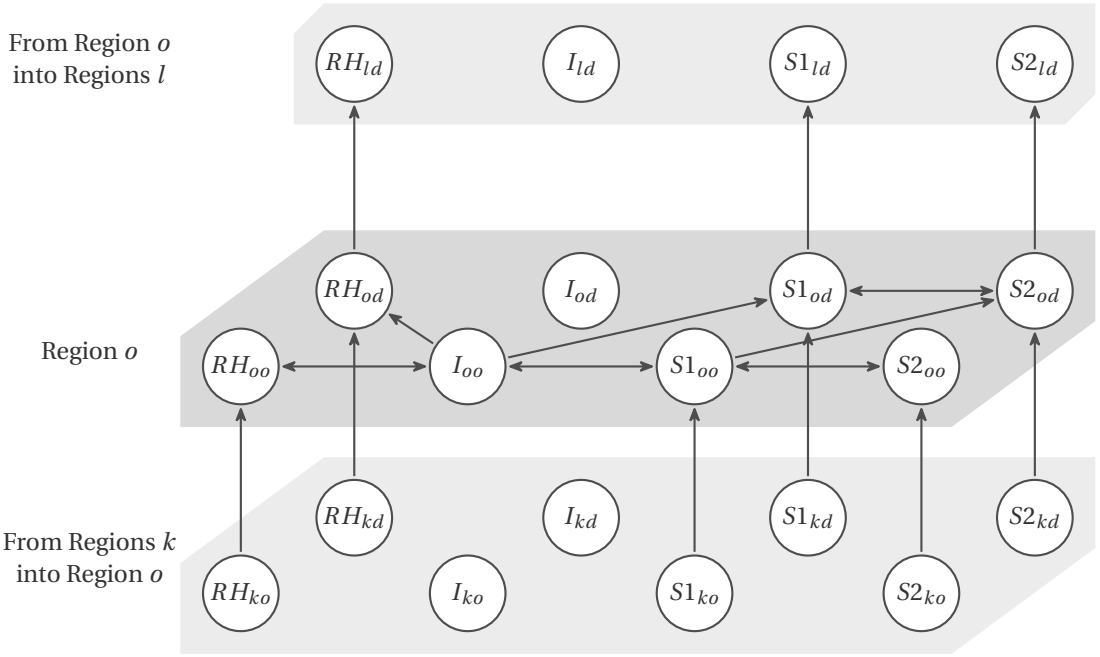


Figure 4.6: General state transition structure focusing on a region o and the inflows and outflows related to this region.

the repositioning path $C^r_{ij^*}$.

$$\dot{\pi}^K_{od}(t) = -\text{Exits} + \text{Entrances} \tag{4.14}$$

Table 4.1: Summary of state transitions in the Markov Chain model.

State	Exits	Entrances
I_{oo}	$\pi^I_{od}(t) \sum_{s\in\mathscr{S}} \sum_{h\in\mathscr{R}} \lambda^s_{oh}(t)$	$\pi^{RH}_{oo}\mu^{RH}_{oo}(t) + \pi^{S1}_{oo}(t)\mu^{S1}_{oo}(t)$
RH_{od}	$\pi^{RH}_{od}(t)\mu^{RH}_{od}(t)$	$\sum_{h\in\mathscr{R}} \pi^I_{oh}(t)\lambda^H_{od}(t) + \sum_{h\in\mathscr{R}_o} \hat{\mu}^{RH}_{hod}(t)$
$S1_{od}$	$\pi^{S1}_{od}(t)\mu^{S1}_{od}(t) + \hat{\lambda}^{S2}_{ohd}(t) + \hat{\lambda}^{S2}_{odh}(t)$	$\sum_{h\in\mathscr{R}} \pi^I_{oh}(t)\lambda^S_{od}(t) + \sum_{h\in\mathscr{R}_o} \hat{\mu}^{S1}_{hod}(t) + \hat{\mu}^{S2}_{od}(t)$
$S2_{od}$	$\pi^{S2}_{od}(t)\mu^{S2}_{od}(t)$	$\sum_{h\in\mathscr{R}} \hat{\lambda}^{S2}_{ohd}(t) + \sum_{h\in\mathscr{R}_o} \hat{\mu}^{S2}_{hod}(t)$

We have to detail the entries of the coefficients for 'Exits' and 'Entrances' in Equation [4.14]. To shorten the description in Table 4.1, we aggregate some particular transitions explained in Equations [4.15]–[4.19]. In particular, Equation [4.15] illustrates that a driver can use different paths on his way to the destination. Equations [4.16] and [4.17] illustrate that shared

ridesplitting drivers (states $S2_{od}$) might complete a ride (Equation [4.16]) before transferring (Equation [4.17]). Finally, Equations [4.18] and [4.19] illustrate that new shared ridesplitting rides can have different delivery orders, such as a *last-in-first-out* (LIFO) order (Equation [4.18]) or a *first-in-first-out* (FIFO) order (Equation [4.19]).

$$\hat{\mu}^{K}_{hod}(t) = \pi^{K}_{hd}(t) \cdot \theta_{hod} \cdot \mu^{K}_{hd}(t) \qquad K \neq S2 \tag{4.15}$$

$$\hat{\mu}^{S2}_{od}(t) = \pi^{S2}_{od}(t) \cdot \vartheta_{ood}(t) \cdot \mu^{S2}_{od}(t) \tag{4.16}$$

$$\hat{\mu}^{S2}_{hod}(t) = \pi^{S2}_{hd}(t) \cdot (1 - \vartheta_{ood}(t)) \cdot \theta_{hod} \cdot \mu^{S2}_{hd}(t) \tag{4.17}$$

$$\hat{\lambda}^{S2}_{ohd}(t) = \pi^{S1}_{od}(t) \cdot \beta^{h}_{od} \cdot \lambda^{S}_{oh}(t) \tag{4.18}$$

$$\hat{\lambda}^{S2}_{odh}(t) = \pi^{S1}_{od}(t) \cdot \beta^{d}_{oh} \cdot \lambda^{S}_{oh}(t) \qquad h \neq d \tag{4.19}$$

Where, $\theta_{hod} \in [0,1]$ distributes transfer flows over its neighboring regions such that the equality $\sum_{h \in \mathscr{R}_o} \theta_{ohd} = 1$ holds; ϑ_{ood} becomes the fraction of shared trips passing through o that will deliver a passenger before continuing to d; and β^{h}_{od} (β^{d}_{oh}) represents the ratio of od (oh) trips that will pass through region h (d).

Deriving state transitions

As described in the previous sections, there are two forms of transition processes, generally represented by $\lambda^{s}_{od}(t)$ and $\mu^{K}_{od}(t)$ for the passenger assignment rate and trip completion/-transfer rates, respectively. Equation [4.20] depicts the assignment rate $\lambda^{s}_{od}(t)$ for service s to a single driver, i.e., the share of all passenger arrival rates perceived by an individual driver. Equation [4.21] depicts the trip completion/transfer rates of a driver, $\mu^{K}_{od}(t)$.[VII] Equation [4.22] calculates the number of available drivers used to compute $\lambda^{s}_{od}(t)$.

[VII] In Beojone and Geroliminis, 2023a, drivers in a state $S1_{od}$ could face interruptions, which would discount the proportion of production not converted into trip completion/transfer rates. Since the CTMC assumes activity time duration is exponential, the average remaining trip length $L^* = (L^2 + \sigma^2)/(2L)$ is equal to the average trip length, thus canceling out the term corresponding to the discount. In summary, a consequence of a memoryless process.

$$\lambda_{od}^{s}(t) = \frac{Q_{od}^{s}(t)}{n_{od}^{av,s}(t)} \tag{4.20}$$

$$\mu_{od}^{K}(t) = \frac{v_o(t)}{L_{od}^{K}(t)} \tag{4.21}$$

$$n_{od}^{av,s}(t) = \begin{cases} n_{od}^{I}(t), & \text{if } s = H \text{ and } o = d \\ n_{oo}^{I}(t) + n_{od}^{I}(t), & \text{if } s = H \text{ and } o \neq d \\ n_{od}^{I}(t) + \sum\limits_{h \in \mathscr{R}\setminus\{d\}} \beta_{od}^{h} n_{oh}^{S1}(t) + \sum\limits_{h \in \mathscr{R}} \beta_{oh}^{d} n_{oh}^{S1}(t) & \text{if } s = S \text{ and } o = d \\ n_{oo}^{I}(t) + n_{od}^{I}(t) + \sum\limits_{h \in \mathscr{R}\setminus\{d\}} \beta_{od}^{h} n_{oh}^{S1}(t) + \sum\limits_{h \in \mathscr{R}} \beta_{oh}^{d} n_{oh}^{S1}(t) & \text{if } s = S \text{ and } o \neq d \end{cases} \tag{4.22}$$

Where $Q_{od}^{s}(t)$ is the total arrival rate of passengers for service s and OD-pair od. $n_{od}^{K}(t)$ stands for the number of drivers currently in state K_{od}. $L_{od}^{K}(t)$ represents the average regional trip length in state K_{od}. These values can be obtained using trip historical data. $v_o(t)$ represents the regional average traveling speed. Recall that the CTMC tracks the activities of a single driver. Therefore, we assume constant $n_{od}^{K}(t)$ during an evaluation to compute $\mu_{od}^{K}(t)$ and $\lambda_{od}^{s}(t)$.

The computation of $n_{od}^{av,s}(t)$ is separated into several cases in Equation [4.22] to account for different matching constraints for different service options. For ride-hailing, the constraints are straightforward, considering idle drivers in the region or repositioning ones with the same od-pair. However, for ridesplitting, the constraints are more complicated. Besides idle and repositioning drivers, one must also account for drivers with a single ridesplitting passenger. Although these drivers are considered for matching, they have additional detour constraints to fulfill. We summarize these constraints by separating two possible cases of matching. In the first case, the driver is considered available if the detours allow him/her to deliver the new passenger in a *first-in-first-out* order, represented as $\sum_{h \in \mathscr{R}\setminus\{d\}} \beta_{od}^{h} n_{oh}^{S1}(t)$. In the second case, the driver is considered available if the detours allow him/her to deliver the new passenger in a *last-in-first-out* order, represented as $\sum_{h \in \mathscr{R}} \beta_{oh}^{d} n_{oh}^{S1}(t)$.

Note that, implicitly in the computation of $\lambda_{od}^{s}(t)$ and particularly $n_{od}^{av,s}(t)$, we assume a constant flow of repositioning drivers. One can interpret it as that the prediction considers that other drivers are likely to keep the same repositioning decisions in the near future (based on the previous information). Therefore, by expansion, it implies that the predictions in the MDCTMC model account for other drivers' decisions in the near future.

In the assumed matching process, one of the passengers will be delivered first in shared ridesplitting. Equation [4.23] uses current demand information to identify the conditional probability $\vartheta_{ood}(t)$ that a shared trip will deliver a passenger in the current region from all the assignable requests resulting in similar od.

$$\vartheta_{ood}(t) = \frac{\sum_{h\in\mathscr{R}} \beta^o_{hd}\left(\lambda^S_{hd}(t) + \lambda^S_{ho}(t)\right)}{\sum_{h\in\mathscr{R}}\sum_{l\in\mathscr{R}} \beta^o_{hl}\beta^l_{od}\left(\beta^l_{hd}\left(\lambda^S_{hl}(t) + \lambda^S_{hd}\right)\right)} \tag{4.23}$$

Where, $\sum_{h\in\mathscr{R}} \beta^o_{hd}\left(\lambda^S_{hd}(t) + \lambda^S_{ho}(t)\right)$ is all demand that would deliver a passenger in region o before continuing to region d. $\sum_{h\in\mathscr{R}}\sum_{l\in\mathscr{R}} \beta^o_{hl}\beta^l_{od}\left(\beta^l_{hd}\left(\lambda^S_{hl}(t) + \lambda^S_{hd}\right)\right)$ indicates all demand for shared ridesplitting trips heading to region d that will pass through o (either for delivering a passenger or just as a passage towards d) before delivering one of the passengers.

Note that the remaining values of θ_{ohd} and β^h_{od} are precomputed parameters to the model. They can be estimated based on historical data on the routes drivers use in their trips.

4.3.2 Estimating drivers' expected revenues

Drivers' earnings come from the fares passengers pay when booking rides, which are composed of (i) a fixed booking fee relative to the reservation of a ride; and (ii) a travel fee relative to the trip distance. The platform keeps a commission for this fare and returns to the drivers the remaining part.

The final price a passenger pays is defined by a fixed fare per booking, called booking fee $f^{s,B}_{od}$, and a variable fare characterized by the trip distance, called traveling fee $f^{s,T}_{od}$. Each of these fees is defined per service and region of origin, and the total fare is $f = f^{s,B}_{od} + f^{s,T}_{od} \cdot D$ (where D is the total trip distance).

The drivers, then, receive $(1-\kappa)\cdot f$ for each served trip fare, where κ is the commission kept by the platform operator. Given the large frequency of events (passenger arrivals, deliveries, etc.), one can approximate the revenue generation for the company by means of a continuous rate. Furthermore, if regions are reasonably homogeneous, drivers' earning generation is approximately uniform and continuous among them. Therefore, a driver receives a part of the booking fees proportionally to the number of received assignments during the evaluation. A driver also earns part of the traveling fees proportional to the kilometers traveled in states with assignments.

To compute the expected revenue for a driver, Equation [4.24] summarizes the expected revenue after commission κ. However, it requires us to decompose the gross revenue into its minor components throughout Equations [4.25] – [4.31]. First, Equation [4.25] separates the expected revenue before the commission into the revenue from booking and traveling fees. Then, Equations [4.26] – [4.28] decompose the booking fees according to the hired service and the respective regional OD-pair until it finally becomes the product of the fixed booking fee in that area and the expected number of booked rides. Particularly, Equation [4.26] decomposes the booking fees from each hired service $s \in \{H, S\}$ (ride-hailing and ridesplitting), and Equation [4.27] further decomposes it to the regional OD-pair $od \in \mathscr{R}^2$. For the revenue

made from traveling fees, Equations [4.29] and [4.31] apply the same logic to achieve the gross revenue from traveling as a product of the fixed traveling fee and the expected passenger-distance traveled.

$$E[R^{\text{net}}] = (1-\kappa)E[R] \tag{4.24}$$

$$E[R] = E[R^B + R^T] = E[R^B] + E[R^T] \tag{4.25}$$

$$E[R^B] = E\left[\sum_s R^{s,B}\right] = \sum_s E\left[R^{s,B}\right] \tag{4.26}$$

$$E[R^{s,B}] = E\left[\sum_{od\in\mathscr{R}} R^{s,B}_{od}\right] = \sum_{od\in\mathscr{R}^2} E\left[R^{s,B}_{od}\right] \tag{4.27}$$

$$E[R^{s,B}_{od}] = E\left[f^{s,B}_{od} n^{s,B}_{od}\right] = f^{s,B}_{od} E\left[n^{s,B}_{od}\right] \tag{4.28}$$

$$E[R^T] = E\left[\sum_s R^{s,T}\right] = \sum_s E\left[R^{s,T}\right] \tag{4.29}$$

$$E[R^{s,T}] = E\left[\sum_{od\in\mathscr{R}^2} R^{s,T}_{od}\right] = \sum_{od\in\mathscr{R}^2} E\left[R^{s,T}_{od}\right] \tag{4.30}$$

$$E[R^{s,T}_{od}] = E\left[f^{s,T}_{od} d^{s,T}_{od}\right] = f^{s,T}_{od} E\left[d^{s,T}_{od}\right] \tag{4.31}$$

Where $n^{s,B}_{od}$ is the number of booked rides for service s from region o to region d; and $d^{s,T}_{od}$ is the passenger-distance traveled for a service s from region o to region d.

Therefore, the remaining unknowns to the expected revenue are the expected number of booked rides and the passenger-distance traveled, which become a function of the time we evaluate. Additionally, each possible decision that a driver can make is assumed to define the starting condition of the forecast, as mentioned in Section 4.3. Therefore, Equations [4.32] and [4.33] estimate these values based on the instantaneous probabilities from the MDCTMC. These estimates are functions of a starting time t_0 and an evaluation period τ for a particular choice γ (remain idle or reposition). Note that, to compute the expected revenue in Equation [4.24], we only have to make the same path of Equations [4.25] – [4.31] backward.

$$E[n^{s,B}_{od}(t_0,\tau|\gamma)] = \int_{t_0}^{t_0+\tau} \lambda^s_{od}(t) \sum_{K\in\mathscr{K}^s_B} \pi^K_{od}(t)\,\mathrm{d}t \qquad \mathscr{K}^s_B = \begin{cases} \{I, RP\}, & \text{if } s = H, \\ \{I, RP, S1\}, & \text{if } s = S \end{cases} \tag{4.32}$$

$$E[d^{s,T}_{od}(t_0,\tau|\gamma)] = \int_{t_0}^{t_0+\tau} v_o(t) \sum_{K\in\mathscr{K}^s_T} n^K_p \pi^K_{od}(t)\,\mathrm{d}t \qquad \mathscr{K}^s_T = \begin{cases} \{RH\}, & \text{if } s = H, \\ \{S1, S2\}, & \text{if } s = S \end{cases} \tag{4.33}$$

Where n^K_p is the number of assigned passengers to a driver in activity K. For states RH and $S1$, $n^K_p = 1$, since they only have a single passenger assigned, while $S2$ have $n^{S2}_p = 2$ because

vehicles carry two passengers in this activity. In reality, drivers do not earn for the distance they drive to pick up a passenger, only for delivering one. Therefore, part of the busy time is not generating revenue. However, we assume that the additional revenue generated by combining pick-up and delivery distances is negligible.

In the proposed setting, the driver is presented with $N \leq |\mathscr{R}|$ possible choices, such that one of these refers to the case where the driver chooses to remain in the current region and the other $N-1$ choices refer to cases where the driver chooses to reposition. Assuming that drivers' objective is to maximize their profit, we expect drivers to follow the estimate indicating the highest revenue. Here, we assume that the cost of repositioning or remaining in the same area is the same, i.e., drivers continue to drive while searching for the next assignment. Therefore, we assume that revenue translates perfectly into perceived utility.

4.4 Computational results

In this prototype application, we represent the central business district of Shenzhen, including parts of the Luohu and Futian Districts. The considered network consists of 1'858 intersections connected by 2'013 road segments. The experiment used a simulator based on Beojone and Geroliminis, 2021b and Beojone and Geroliminis, 2023a, using the Floyd-Warshall algorithm to compute shortest paths and a Speed-MFD to estimate average traveling speeds. A network-weighted k-mean algorithm separated the area into three distinct regions and the Speed-MFD data (Figure 4.7). Detailed clustering algorithms can separate the region based on traffic data, such as Saedi et al., 2020, Saeedmanesh and Geroliminis, 2016 and Hans et al., 2014.

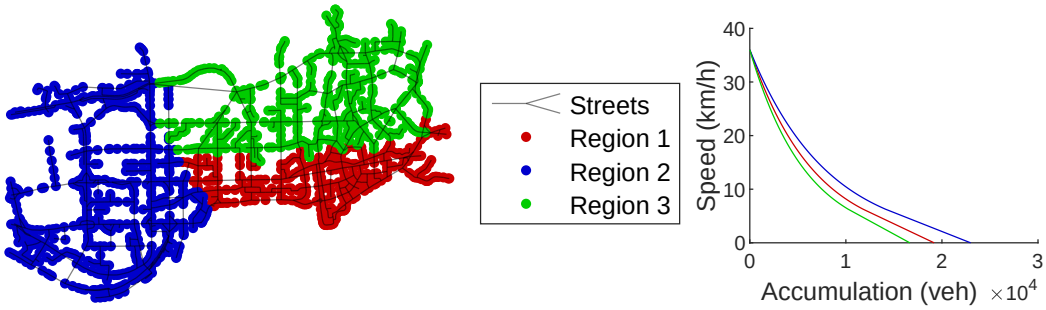


Figure 4.7: Settings of the experiments. (Left) urban area map and regions with respective centroids of the k-mean problems. (Right) Regional speed-MFD.

We assume a non-homogeneous Poisson arrival process for all travelers in the area to highlight the effects of imbalanced demand. In Figure 4.8, we illustrate the total arrival rates in each regional OD-pair, indicating an hourly change in the region originating the most trips. Note that around 85% of arriving travelers use private vehicles, while the remaining use one of the ride-sourcing service options.

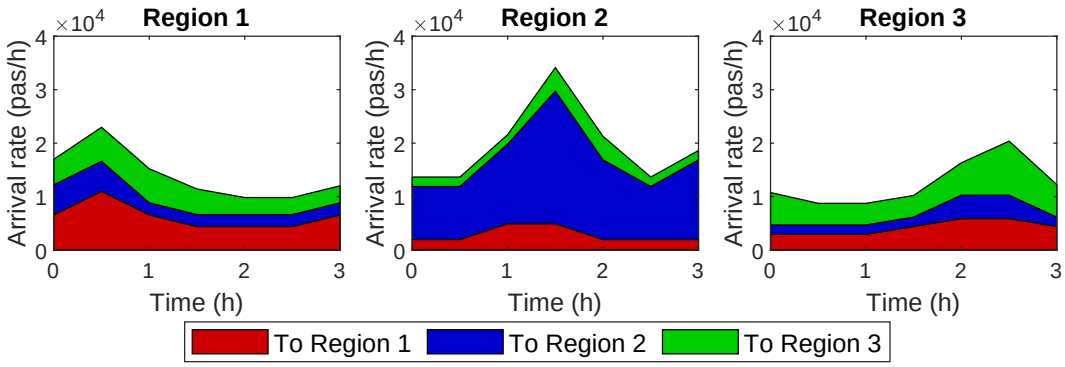


Figure 4.8: Total arrival rates grouped by region of origin.

4.4.1 Sensitivity analysis and benchmark comparison

The first analysis of the repositioning strategy focuses on the sensitivity to the fraction of drivers receiving the repositioning guidance. One can correlate such scenarios with the operator selecting a group of drivers for a 'loyalty program,' and, as part of the benefits, these drivers receive improved guidance in their search for assignments and increased revenues. Note that the selection criteria used by TNCs to include drivers in such programs are outside the scope of this paper. Instead, the focus is on the effects of potential selection processes that include a ratio of X% of all the drivers in the operating fleet. That said, we refer to drivers who receive guidance as 'guided' ones, whereas we refer to the others as 'unguided.'

As argued earlier, the guidance is a suggestion of the decision that would maximize a driver's revenues in the near-future. Therefore, we assume a logit decision process, where the utility of each option is depicted exclusively by the revenue it generates to a 'guided' driver.[VIII] Finally, we define that 'unguided' drivers look for the region with the highest demand per driver rate (similar to a 'high demand' flag in current ride-sourcing operations).

Additionally, we evaluated the results for fleet sizes of 2000, 2500, and 3000 active drivers in ride-sourcing services. We ran cases with 0%, 25%, 50%, 75%, and 100% of drivers covered in the 'loyalty program.' In total, we tested 5 instances of each scenario before the evaluation. As final parameters, we considered the same fares from Beojone and Geroliminis, 2021b, where booking fares were US$2.20 and US$2.00, and traveling fares were US$1.00 and US$0.80 per kilometer for ride-hailing and ridesplitting, respectively.

As argued earlier, other repositioning strategies exist to improve service quality, but they fall short on their reactive nature and/or full compliance assumptions. In this section, we compare the proposed strategy with other benchmark strategies regarding improvements in service quality. In addition to a 'No-repositioning' base case (drivers always remain in their current

[VIII] The logit model is described as a Multinomial logit model, where the choice set is composed of all regions (including the current one, as a choice for not repositioning) and their utilities are represented by the expected revenue a driver will obtain if deciding to move to the respective region. Therefore, the probability that an individual chooses the alternative r (region $r \in \mathscr{R}$) is $P(r|\mathscr{R}) = \exp(E[R_r^{\text{net}}]) / \sum_{o \in \mathscr{R}} \exp(E[R_o^{\text{net}}])$.

area when becoming idle), we evaluate the strategies named below:

1. 'Proposed': proposed strategy described in Sections 4.2–4.3.2.
2. 'Past-revenue': Provides a portion of 'guided' drivers with the average revenue drivers made in the past τ hours for each area (same horizon as the prediction horizon, but for past events).
3. 'Past-loss': Dispatches the closest idle driver to the area of a recently lost request, similar to Alonso-Mora et al., 2017.
4. 'Coverage': Performs optimal coverage control, distributing all idle drivers according to the demand distribution in the area (Zhu et al., 2022). It computes idle vehicles' coverage based on a Voronoi tessellation, which partitions the area into responsibility zones for each vehicle.

It is important to highlight that the 'Past-revenue' strategy has a reactive nature but allows drivers to decide, whereas the 'Past-loss' strategy is reactive and assumes full compliance with the instructions. On the other hand, as pointed out earlier in the text, the 'Coverage' strategy tries to position drivers before the requests arrive but assumes that the operator has control of a fully compliant fleet that is easier to implement in a scenario with autonomous vehicles. We must note that, in Zhu et al., 2022, the problem is optimized for a static geographical distribution of demand, i.e., the spatial imbalance is constant over time, which differs from the tested case. To cope with this difference, we estimated the geographical distribution of demand for the near-future based on the arrival data and the same forecast horizon of the proposed method.

The foremost objective of any repositioning strategy is to improve service quality, especially by making the service available in previously uncovered areas. Figure 4.9 compares the number of unattended service requests (abandonments) for all evaluated strategies. Firstly, when no drivers receive guidance and base their decisions on 'high-demand areas' information (0% guidance), service is worsened by increasing abandonments compared to the base case. Then, reactive strategies ('Past revenues' and 'Past losses') performed poorly, with little to no improvement compared to the base case. The case using 'Past losses' could only show some improvement at a smaller fleet size of 2000 vehicles (a decrease of 13% in abandonments). The cases using 'Past revenues' had a decrease only at larger fleet sizes with comparable results to the 'Coverage' method when guiding between 50% and 75% of drivers. The most intriguing comparison is between the 'Proposed' and the 'Coverage' approaches. Although optimized, the 'Coverage' approach was outperformed by the 'Proposed' approach when 50% (or more in larger fleet sizes) of drivers receive guidance. One possible explanation to this is on the difficulty that the 'Coverage' approach faces when multiple vehicles are very close to each other (sharing the same nearest intersection), which forces the Voronoi diagram to consider these vehicles as a single one and provide instructions only to one of them, while

the remaining vehicles remain in the positions. Another limitation of the 'Coverage' approach is that it relies on the historical demand spatial distribution, which might not capture how the distribution changes over the near-future. Therefore, the 'Proposed' method was the most successful in serving previously unattended requests by providing half of the drivers with repositioning guidance, decreasing abandonments by 61% for a fleet of 2000 vehicles. Nevertheless, we must note that there are little increases in abandonments if the guidance is available for more than 50% of the drivers. In Section 4.4.3, we explore the limitations causing this increase in abandonments, being either the revenue-based decision process (Section 4.3) or the path guidance definition (Section 4.2). A promising result is that a fleet of 2000 vehicles with 50% guided drivers has almost identical performance with system of 3000 vehicles with no guidance.

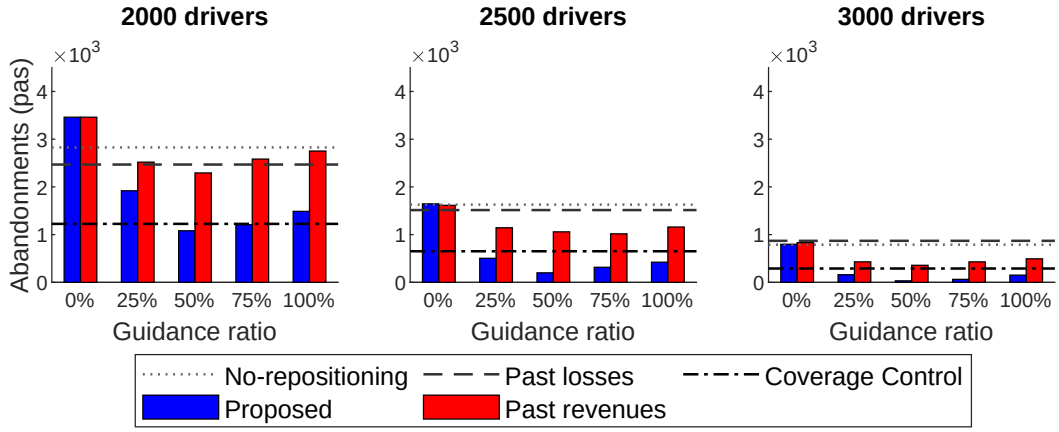


Figure 4.9: Summary of passenger abandonment results for all compared strategies.

So far, we have compared the abandonments associated with each strategy without investigating the causes of these results. Therefore, in Figure 4.10, we present results associated with the vehicles themselves for a fleet of 2000 vehicles and 50% of guidance ratio in the 'Proposed' and 'Past-revenues' approaches. It highlights vehicles' response to each repositioning strategy in terms of the number of repositioning vehicles and the VKT for repositioning activities. Firstly, the 'Past losses' strategy moves almost no vehicles, peaking at 7 simultaneous drivers, therefore, we do not include it in the coming analysis. Then, a lower plateau where the 'Coverage' strategy remains between 50-150 simultaneous vehicles alongside the 'guided' drivers in the 'Proposed' approach. Note that the numbers in the 'Coverage' strategy present some significant noise due to the extremely short (if any) repositioning instructions provided to all drivers. Finally, 'unguided' drivers in the 'Proposed' approach and all drivers in the 'Past revenues' approach form a higher plateau (mostly between 250-300 simultaneous vehicles). The direct consequence of these numbers is seen in the VKT associated with repositioning activities, where these vehicles travel more than 1.5×10^4km only for repositioning activities. On the other hand, 'guided' vehicles in the 'Proposed' 'Coverage' approach traveled less than one-third of this distance. The previous, combined with the information on the number of abandonments, highlights that relocating too many vehicles decreases the system service

capacity.

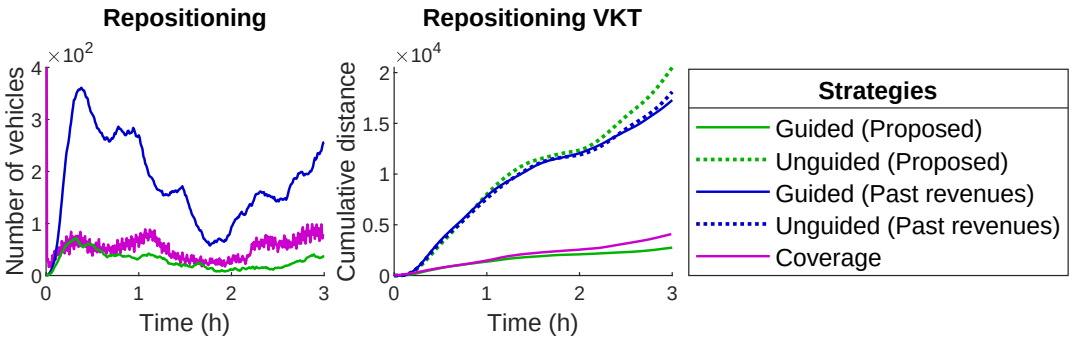


Figure 4.10: Comparative among all tested benchmarks for (left) the number of repositioning vehicles, and (right) VKT associated with repositioning activities.

Analyzing the compared occupancy ratio in Figure 4.11, all strategies increase the time vehicles spend occupied. However, the same groups that reposition the most have valleys in the occupancy rate, being momentarily less occupied than the base case. The 'guided' drivers in the 'Proposed' approach present the highest occupancy rate, being 10% of the time more occupied than the base case. We must highlight that 'unguided' drivers in the 'Proposed' approach had higher occupancy around 1.5h of the simulation, coinciding with the period that they have the fewest vehicles repositioning and where the VKT slope flattens. It happens because this is the period where there is a peak in demand for Region 2, the area with the largest demand (see Figure 4.8), thus overlapping the 'high-demand area' flag with the actual area where vehicles are required.

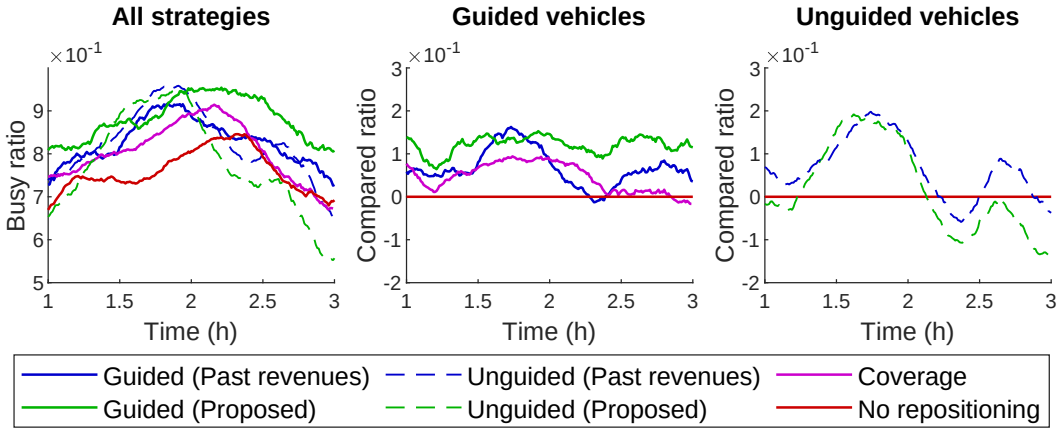


Figure 4.11: Fraction of busy vehicles (occupancy ratio) for scenarios with 2000 ride-sourcing drivers and 50% guidance ratio. (Left) Fraction of busy guided vehicles. (Center) Difference of the occupancy ratios of 'guided' drivers to the 'No repositioning' case. (Right) Difference of the occupancy ratios for 'unguided' drivers.

Combining the dynamics of drivers' activities and abandonments can provide further insights into how each repositioning strategy affects service quality. Figure 4.12 illustrates these dy-

namics for four cases with 2000 vehicles. Firstly, in the 'No repositioning' case, abandonments peak when the demand peaks in different regions, highlighting the imbalance of the demand and supply of drivers. Interestingly, these are also periods when the number of idle vehicles reaches local maxima. Secondly, given the delays in the provided revenues of the 'Past revenues' strategy, abandonments take longer to decrease and never return to null numbers. However, the peak in abandonments has a significant decrease. Thirdly, the 'Coverage' strategy also had a lowered peak in abandonments but slightly shifted to the right and was followed by null values after 0.5 hours. Finally, the 'Proposed' strategy had an even lower peak in abandonments, which was also shifted to the right, followed by a decrease until reaching null numbers. The number of busy vehicles is maximized before 2 hours in the 'Past revenues' and 'Proposed' strategies, whereas in the 'No repositioning' and 'Coverage' are slightly shifted to some minutes later.

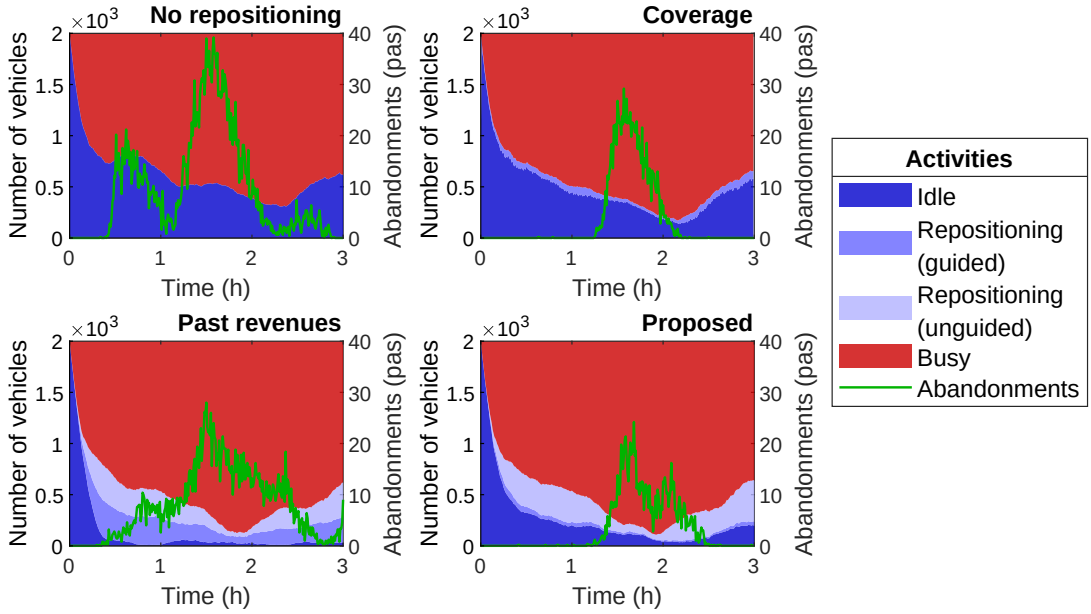


Figure 4.12: Dynamic activities of ride-sourcing vehicles and number of lost requests in the same period (intervals of 36 seconds) for scenarios with 2000 ride-sourcing drivers and 50% guidance ratio.

Most attention to the ride-sourcing effect over congestion goes to their deadheading. We consider the distances traveled for pick-up activities as deadheading because the evaluated service does not charge passengers for the pick-up distance, only for the booking and the delivery distance. In Figure 4.13, we summarize the deadheading accumulated for each strategy for the same tested cases with 2000 vehicles and a 50% guidance ratio. The only strategy to increase it, both in the unassigned and pick-up activities, is the 'Past-revenues' strategy, illustrating a situation where passengers could be served but had to wait longer for a driver. In the 'Proposed' approach, the unassigned deadheading is minimized, but in the 'Coverage' approach, the 'Pick-up' deadheading is minimized. It highlights two distinct points about these strategies. First, while the 'Proposed' approach maximizes the chances of a driver

being assigned, the driver only needs to be as close as the passengers' waiting time tolerance accepts. Second, the 'Coverage' approach has minimal 'Pick-up' deadheading because it mirrors the expected demand distribution, thus, bringing drivers close to the area where the next arriving passengers are expected to come. However, since only one vehicle is enough to consider the area covered, once a vehicle is assigned a new passenger, another vehicle must move to that area to cover it, which might not be close and require some time to move.

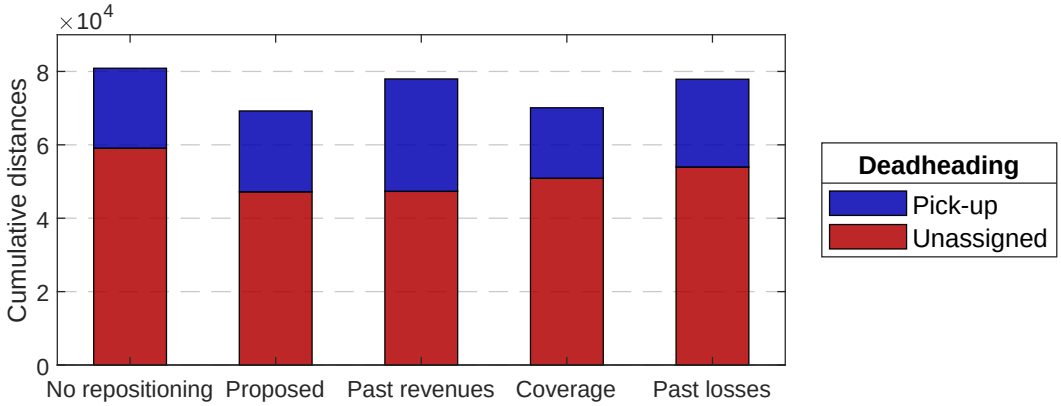


Figure 4.13: Summary of the deadheading associated with each repositioning strategy, separated into unassigned and pick-up kilometrages. Scenario with 2000 drivers and 50% of guidance ratio in the 'Proposed' and 'Past revenues' methods.

Before we evaluate the ability of the 'Proposed' strategy to retain drivers, it is interesting to assess how guided drivers adhered to the repositioning suggestions. Figure 4.14 illustrates this as the fraction of times that guided drivers followed the decision with the highest forecasted revenue. At an initial glance, guided drivers' compliance (following the highest revenue forecast) decreases with the number of drivers receiving guidance, both by expanding the guidance ratio and enlarging the operating fleet size. In that regard, the 'Proposed' strategy shows lower sensitivity when compared to the 'Past revenues' strategy. On the one hand, compliance in the 'Proposed' strategy ranges from 78% to 91%, while it ranges between 66% and 92% in the 'Past revenues' strategy. In both cases, the sensitivity to the size of the operating fleets is higher than the guidance ratio (in terms of the number of guided drivers).

4.4.2 Revenue forecasting scheme evaluation

Since the proposed repositioning framework must persuade 'guided' drivers, the revenue generated should be higher than those of 'unguided' drivers. Otherwise, the framework would lose attractiveness, losing compliance in the long term. Hence, as a first step to evaluate the proposed repositioning framework, we compared the revenues made by each group of drivers. Figure 4.15 shows the average revenues of both groups of drivers at different guidance ratios (fraction of the fleet that receives repositioning guidance information). Firstly, average revenues increased compared to a scenario where drivers never relocate in all cases with guidance (the exception occurs at 0% guidance ratio). As one could expect, large fleet sizes

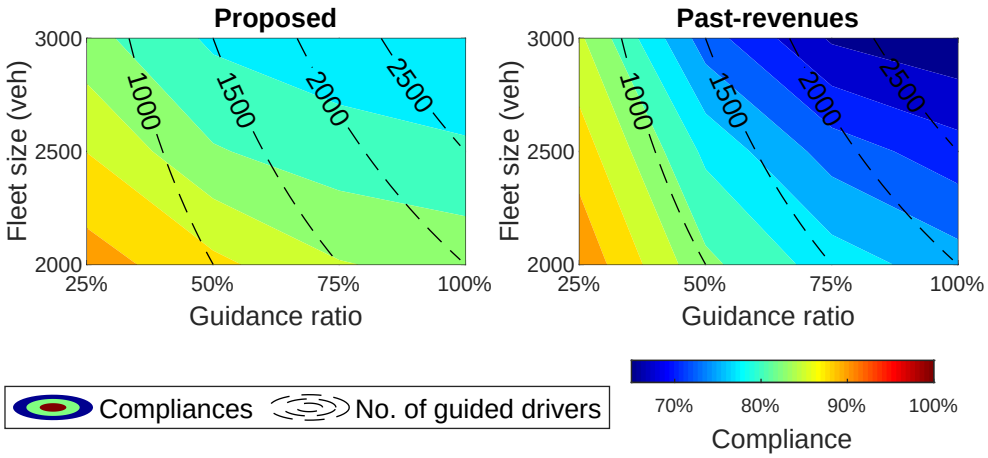


Figure 4.14: Fraction of guided drivers taking the decision predicted to maximize his/her revenues. Comparison between the 'Proposed' and the 'Past revenues' strategies.

also decrease the average revenue and the effects of guidance, indicating that most demand is covered by the additional vehicles with no need for relocating drivers. We must point out that guided drivers consistently have higher revenues than non-compliant ones. It highlights that the proposed framework captures the possibility of areas with lower demand being more profitable. However, it is interesting to observe that as the operator expands the number of 'guided' drivers (more than 75% of the fleet), the combined average revenue slightly decreases.

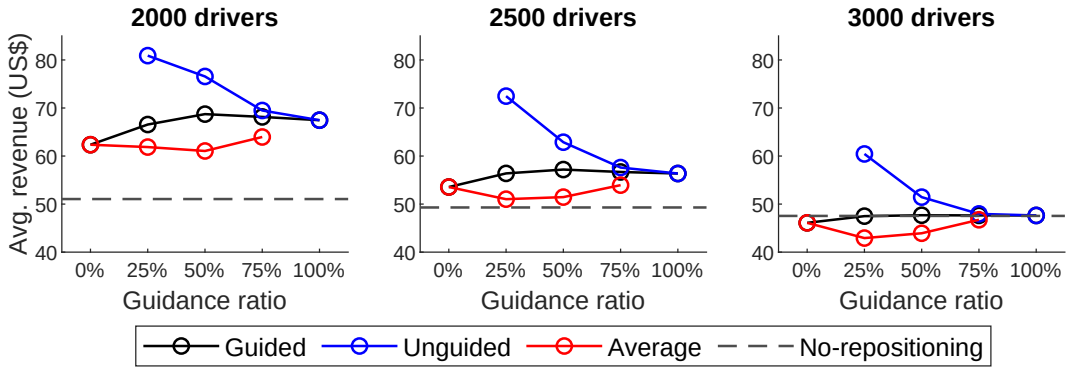


Figure 4.15: Average revenues of 'guided' and 'unguided' drivers following by a combination of these in scenarios with different penetration rates of the loyalty program, compared to the base 'No-repositioning' case.

Besides the average revenue, it is interesting to take a closer look at individual driver revenues and understand how these are distributed. Figure 4.16 shows the histograms (normalized as probability density functions) of the revenues for 'unguided' and 'guided' drivers in the base 'No-repositioning' case, one with 0% guidance ratio and one with 25% guidance ratio for a service fleet of 2000 drivers. With a left-skewed distribution, the 'No repositioning' scenario has a lower average revenue average, just slightly higher than the average for 'unguided'

drivers in the compared scenarios. In the 0% guidance ratio, the left tail of the distribution disappears, leading to a symmetric distribution of revenues with the lowest average from the other cases. One can observe that the distribution of revenues was unimodal, with the average close to the mode (scenarios with 0% and 25% of 'guided' drivers), creating a clear distinction between these groups. The standard deviation is slightly smaller for 'guided' drivers (US$8.96 *vs* US$10.14). Additionally, the average revenue of compliant drivers was higher than the revenue for 97% of non-compliant ones (85% if observing one standard deviation behind the average).

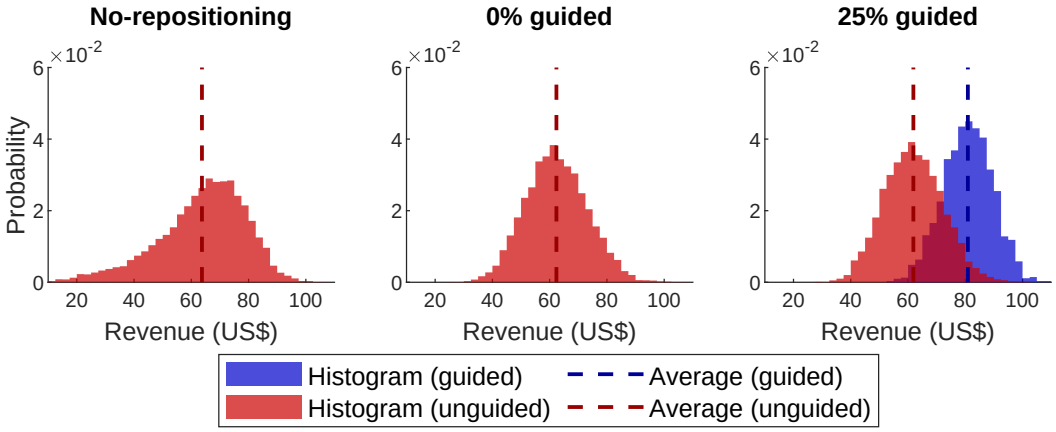


Figure 4.16: Histograms of revenues in three separated instances with 2000 drivers.

To complete the analysis of the drivers' choices, we must evaluate the revenue prediction process. Figure 4.17 illustrates the repositioning options for two drivers near the beginning of the simulation (around 0.15h from the start). Note that the figure only indicates the forecast time instead of the simulation time, and it is as long as the prediction horizon $\tau = 0.5$h. Recall that, around this period, most demand originates from Region 1, followed closely by Region 2 (see Figure 4.8). At this moment, more than 1000 drivers are busy, while 400 and 270 are idle in Regions 1 and 2, respectively. In the first example, the driver is far from the border of other regions, thus showing that it has a higher chance of getting higher revenue in the current region instead of losing capacity by moving to another one. In the second example, however, the driver is closer to another region, which makes the expected revenue for staying in Region 2 or moving to Region 3 comparable and likely options. In both examples, Region 1 is farther from the driver's position, thus, becoming an inferior option. We must highlight how the MDCTMC model captures the time required for moving among different regions, depending on the drivers' initial position, significantly changing the slope of the revenue forecasts for shorter or longer periods, directly affecting the final result and the driver's decision.

A few last measurements are necessary to understand the implications of focusing on possible assignments on the repositioning path instead of the destination only in Section 4.2. Table 4.2 presents some aggregated statistics regarding the assignments and repositioning activities in the scenario with 2000 drivers and a 50% guidance ratio. Note that we are comparing the numbers of assignments, which, for ridesplitting, will differ from the number of passengers

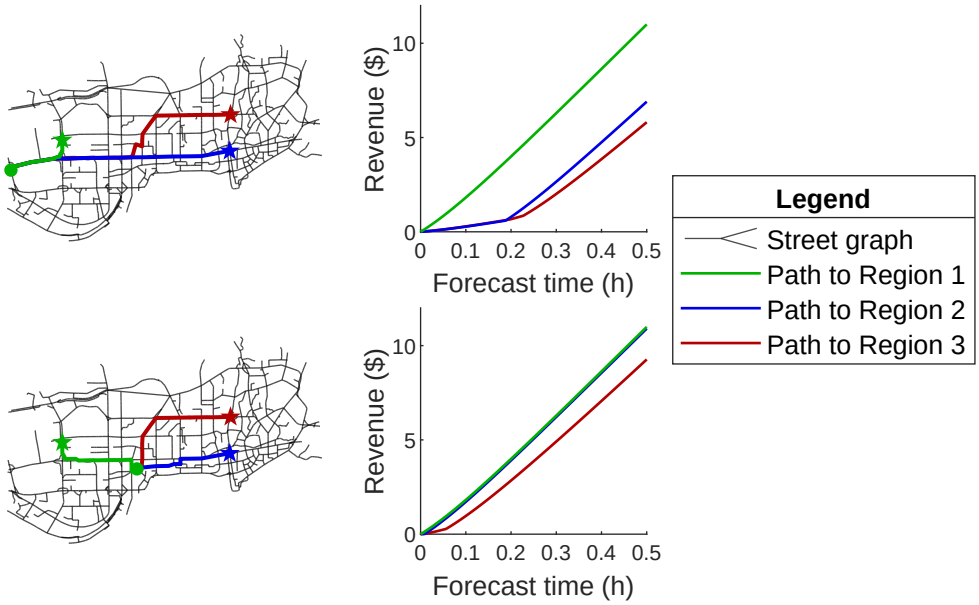


Figure 4.17: Two examples of revenue forecasts in different scenarios with the suggested destinations and paths.

who share in-vehicle space. Given the FCFS matching, the assignment depends on the driver's current activities. For instance, a driver always starts a new ridesplitting trip with a single assignment; all the subsequent assignments are going to be shared assignments until the driver becomes idle again. Firstly, we checked that nearly 1 out of 4 assignments (23%) to guided drivers occur while they are repositioning. In a deeper look, it showed, however, that only 46% of these assignments are for ridesplitting when nearly 60% of all requests (and assignments) hired this service. To understand this, we saw that nearly half (49%) of all ridesplitting assignments are classified as shared trips. On the one hand, these results illustrate a significant number of assignments during repositioning. On the other hand, they show that the drivers in service are usually closer to incoming ridesplitting requests than the repositioning ones. This observation illustrates that alternatives to the shortest path could improve matching possibilities for ridesplitting requests, further increasing the chances of matching during repositioning activities.

4.4.3 Impact of drivers choices

As argued all along the paper, drivers in ride-sourcing services perform a series of decisions, including the decisions regarding repositioning evaluated in the proposed repositioning strategy. These decisions were assumed to follow a logit process.[IX] In order to evaluate the impact of drivers' free will, we explore the effect of replacing this assumption by applying two

[IX] See Footnote VIII.

Table 4.2: Other assignment statistics related to repositioning in a scenario with 2000 drivers and 50% guidance ratio.

Statistics	Outcomes	
Frequency guided drivers were assigned while repositioning	Repositioning (23%)	Not repositioning (77%)
Services hired for while repositioning	Ride-hailing (54%)	Ridesplitting (46%)
Overall hired service distribution	Ride-hailing (39%)	Ridesplitting (61%)
Frequency of shared ridesplitting assignments	Shared (49%)	Single (51%)

deterministic processes named below while we refer to the original process as 'Logit-based.'

- Deterministic based on the revenue forecast (called 'Det. revenue'): the decision follows the instruction with the highest expected revenue (value $E[R^{\text{net}}]$, as computed in Section 4.3.2).
- Deterministic based on the expected number of requests (called 'Det. requests'): the decision follows the instruction with the highest expected number of assignments (value Π_i, as computed in Section 4.2).

Additionally, by separating these cases according to the major elements in the proposed approach, we shed some light on their individual undetected limitations and bottlenecks for future implementations.

As previously shown in Figure 4.9, abandonments could slightly rise at increased guidance ratios. In Figure 4.18, passenger abandonments are explored for the described decision processes ('Logit-based', 'Det. revenue', and 'Det. requests') and different fleet sizes. The general decreasing trend in abandonments is kept in terms of the guidance ratio. All processes had results close to the original ones in the 'Logit-based' with no significant difference. The only exception occurs for the 'Det. requests' process with 2000 drivers at 100% guidance ratio, which has significantly lower abandonments than the other strategies. The previous highlights that it could provide solutions to improving the number of assignments, despite the approximations in Section 4.2 regarding demand coverage and service.

From the previous findings, one can expect that the differences in abandonments will be converted into drivers' revenues. Figure 4.19 explores driver revenues for the described decision processes. Firstly, we must point out that the average revenues for all drivers did not change for the different processes. However, even with the lowest abandonment numbers, the 'Det. requests' process had little impact on the average revenue (only about US$1.50 higher at a 100% guidance ratio for 2000 vehicles). As a general result, since the 'Det. requests' process does not focus directly on increasing revenues, guided drivers through this process have lower average revenues than those under revenue-based process in the same scenarios.

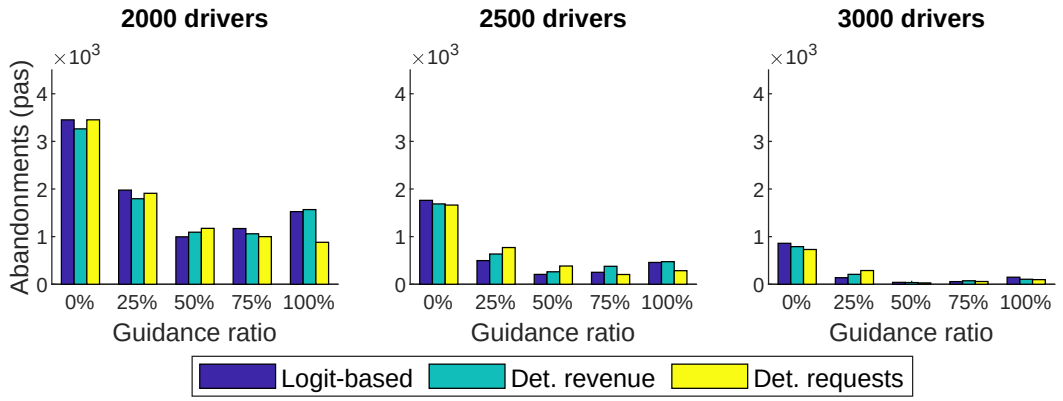


Figure 4.18: Comparison of passenger abandonments in the proposed strategy with deterministic repositioning response.

For instance, guided drivers at a 25% guidance ratio and a fleet size of 2000 vehicles made an average of US$76.00, while they made around US$81.00 in revenue-based processes. The previous highlights the limitations of the assumption that increasing the number of served requests increases revenues in the path orientation process. At the same time, unguided drivers had higher revenues compared to revenue-based approaches, decreasing the average revenue gap between guided and unguided drivers while maintaining the same global average. The 'Det. revenue' process only differentiates itself from the 'Logit-based' one at larger fleet sizes when it presents an average revenue of US$3.00 higher at a 25% guidance ratio.

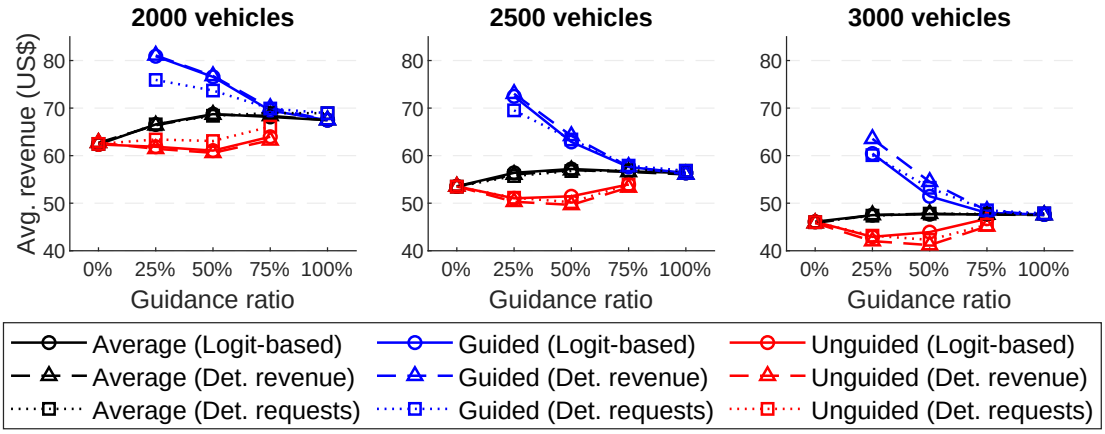


Figure 4.19: Comparison of average driver revenues in the proposed strategy with deterministic repositioning response.

From the company's perspective, these results should translate into its revenues for the period. Table 4.3 summarizes the profit of the company based on the commissions it charges from each served ride. In general, the company's revenues are not too sensitive to the deterministic repositioning response. With the exception that the 'Det. requests' strategy achieves higher revenues (beyond the margin of error for the simulations, which ranges mostly around

US$100.00) at a guidance ratio of 100% for a fleet of 2000 vehicles. At larger fleet sizes (2500 and 3000 vehicles), the revenues do not vary significantly to the guidance ratios of 25% or higher. All the previous show that applying the 'Proposed' repositioning strategy increases the platform revenue from commissions. At the same time, these results show little to no effect on macroscopic measurements for assuming logit-based decisions compared to deterministic (dictatorship) scenarios in repositioning response.

Table 4.3: Comparison of the total revenue of the service provider in the proposed strategy with deterministic repositioning response. Revenues in US$.

Guidance ratio		0%	25%	50%	75%	100%
2000 vehicles	Logit-based	22'009	23'492	24'256	24'051	23'810
	Det. revenue	22'132	23'430	24'224	24'126	23'813
	Det. requests	22'046	23'497	24'136	24'336	24'326
2500 vehicles	Logit-based	23'626	24'876	25'230	25'008	24'866
	Det. revenue	23'612	24'715	25'125	24'960	24'728
	Det. requests	23'590	24'580	25'041	25'100	25'066
3000 vehicles	Logit-based	24'404	25'142	25'244	25'223	25'219
	Det. revenue	24'257	25'157	25'304	25'207	25'149
	Det. requests	24'357	25'079	25'308	25'273	25'334

4.5 Summary

In this chapter, we evaluate the potential repositioning response of drivers when provided guidance based on estimates of their earnings in a system offering ride-hailing (solo) and ridesplitting (shared) rides. Therefore, we develop a strategy that enumerates the best regional repositioning destination based on the expected number of requests a driver will encounter during the forecast horizon. A mixed continuous-discrete time Markov Chain (MDCTMC) is developed to predict a driver's activities and the associated revenues. In summary, the developed strategy provides a group of drivers with individualized near-future revenue estimates guiding drivers toward repositioning decisions that are more likely to maximize their earnings. Our main findings indicate that if the operator selects only a fraction of active drivers to provide guidance, these are likely to expect higher earnings than those without guidance. We also show that it manages to decrease the number of unserved requests compared to several state-of-art benchmarks while increasing vehicle occupancy and decreasing the deadheading.

5 A hierarchical control framework for vehicle repositioning in ride-hailing systems

This chapter is based on the following extended abstract:

- Beojone, C. V., Zhu, P., Sırmatel, I. İ., & Geroliminis, N. (2023). A hierarchical control framework for vehicle repositioning in ride-hailing systems. *25th International Symposium on Transportation and Traffic Theory (accepted for full paper submission)*

It must be highlighted also that this is not a complete study in this version of the thesis. Since this is a collaborative effort, this chapter expands the content related to the middle-layer in comparison with the original extended abstract, where the candidate's contributions are more pronounced.

5.1 Introduction

Spatiotemporal variations in demand can create supply imbalances between drivers and passengers in on-demand mobility services, manifesting as deterioration of system efficiency and service quality. Therefore, it is expected that an efficient fleet management strategy, that keeps drivers well-distributed in space and time over the served area, is of critical importance to provide satisfying mobility service.

In this chapter, we propose a hierarchical control strategy for the relocation of idle ride-sourcing vehicles, for addressing the gap between proactive repositioning strategies and micro-management of vehicles in such activities. The upper-layer utilizes an aggregated model, which is an approximation of trip-based MFD modeling approach (building on Beojone and Geroliminis, 2023a). A model predictive control (MPC) framework is employed to determine the number of idle vehicles to be relocated for each pair of regions. Unlike perimeter control MPC methods, fleet management MPC requires the integration of more sophisticated MFD-

based models describing mixed dynamics of private vehicles and taxis. In the lower-layer, given the demand density over the current region, a coverage control scheme operates to distribute the vehicles within the region to achieve a demand-aligned configuration, which provides each vehicle with relatively detailed (i.e., intersection/node-level) position guidance. To bridge both layers, a middle-layer mechanism is developed for converting the upper-layer decisions into dispatching commands for individual vehicles by solving an Assignment Problem, which minimizes the distance required to achieve the optimal coverage and repositioning decisions. An agent-based simulator built on a trip-based MFD model is utilized with empirical taxi data from a real network of Shenzhen for validation of the proposed strategy. Compared with control strategies using exclusively the upper-layer or the lower layer, the results indicate that the proposed hierarchical framework method yields performance improvements by answering more requests with lower waiting times. Note that, differently from the previous Chapter 4, the intent is to identify the most optimized repositioning strategy, assuming vehicles full compliance with the provided instructions (i.e., autonomous vehicles operated by the service provider), creating a best case scenario for repositioning decision benchmarks.

Following the motivation and the challenges raised regarding repositioning ride-sourcing vehicles with minimal human influence, which is given in Section 1.2.4 of Chapter 1, the remainder of the chapter has the following structure. Section 5.2 presents the proposed hierarchical control framework and the details of the methodology employed in each layer. Section 5.3 depicts numerical results of the entire proposed framework compared with the results for each of its layers and a brief evaluation of the repositioning decisions.

5.2 Hierarchical control framework for vehicle repositioning

Implementing a controller for a large-scale system, one may face problems such as high computational effort due to complex models and high dimensions required for accurate network modeling, especially if the model and controller are developed to compute control actions for every individual vehicle over the whole network. One way to solve this problem is to build a hierarchical control structure. Such structures decompose the control problem into a hierarchy of decision-making levels, and operate via coordinating between the actions of an upper-layer controller (operating at the aggregated traffic level) and a lower-layer controller (managing individual vehicles). The control structure is shown in Figure 5.1.

The upper-layer controller collects aggregated information, such as how many empty vehicles are in each region, from all urban regions at a relatively large update period T_u. The control action generated from the upper-layer determines how many vehicles should stay in current regions and how many vehicles should relocate to other regions, in order to improve availability and thus minimize the total waiting time of passengers. Furthermore, the middle-layer transfers the obtained upper-layer guidance to the lower-layer and specifies which vehicle should stay or move, considering the travel costs caused by repositioning. It is operated within each region and requires relatively more detailed information, such as the coordinates of each

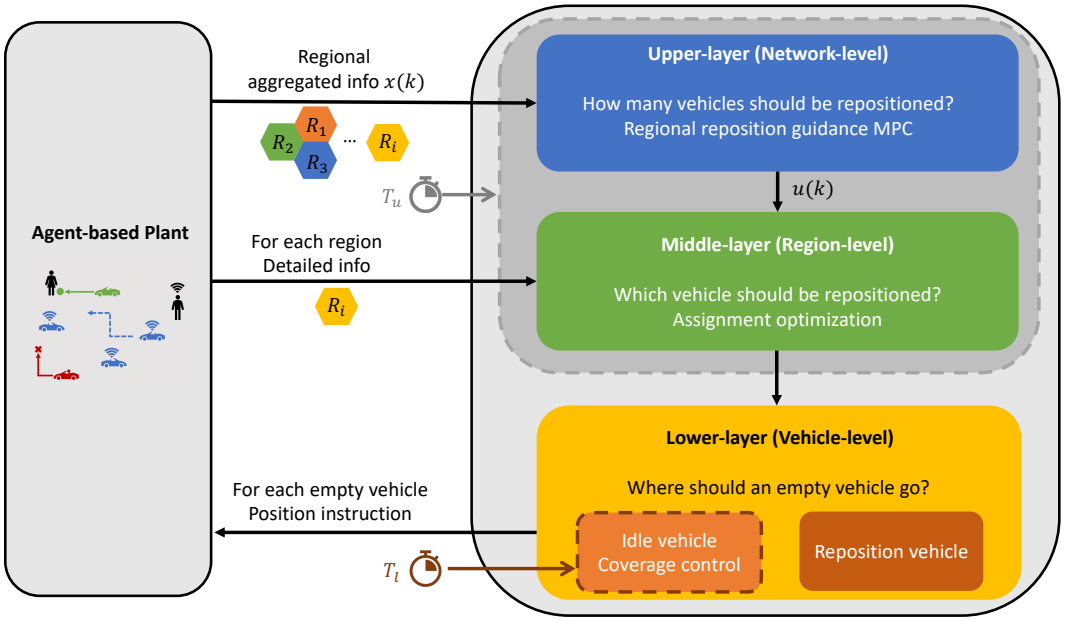


Figure 5.1: A hierarchical control framework for vehicle rebalancing.

vehicle and whether it is occupied or not. Note that the middle-layer can only be activated when the upper-layer is active. The lower-layer is operated in a distributed manner so that each vehicle can obtain its own control action, which facilitates its implementation at a fast update period T_l. The empty vehicles that are commanded to stay in the current region (i.e., idle vehicles, see the left part of lower-layer in Figure 5.1) communicate and cooperate with each other to achieve better vehicle position configuration, while the rest of the vehicles (i.e., repositioning vehicles, see the right part of the lower-layer in Figure 5.1) are be guided to other desired regions as per the relocation commands.

5.2.1 Upper-layer: Prediction model and MPC

The designed framework incorporates urban traffic dynamics, describing the macroscopic behavior of private (i.e., non-taxi) vehicles, which forms the major part of overall urban traffic. In such a case, we can classify the vehicles in the system based on their ongoing activities and regional movements: vacant vehicles, consisting of idle and repositioning vehicles, are denoted as V, busy vehicles and private vehicles are denoted by B and P, respectively. Note that vehicles in states B are assumed completely busy and cannot receive new assignments. At the same time, vehicles in states V are considered available for new assignments when idling in one region. While a more general formulation will be presented in the full paper, a two-region example is given here.

The developed dynamical model summarizes past travel behavior into total remaining distance states (Sirmatel et al., 2021). The model also captures movement of ride-hailing drivers between the urban regions. Let $K \in \{V, B, P\}$ and assume the urban area is partitioned into

R_n regions, $\mathscr{R} = \{1,2\}$. Equations [5.1] and [5.2] summarize the basic structure of the model at each state, as previously proposed in Beojone and Geroliminis, 2023a. Considering the remaining distance to be traveled effects, total trip completion/transfer terms $O^K_{od}(t)$ (from region o to d) can be written as in Equation [5.3].

$$\dot{n}^K_{od}(t) = I^K_{od}(t) - O^K_{od}(t), \tag{5.1}$$

$$\dot{m}^K_{od}(t) = I^K_{od}(t) L^K_{od}(t) - n^K_{od}(t) v_o(t), \tag{5.2}$$

$$O^K_{od}(t) = \left[n^K_{od}(t) + \alpha_{od} \left(\frac{m^K_{od}(t)}{L^*_{K_{od}}} - n^K_{od}(t) \right) \right] \frac{v_o(t)}{L^K_{od}(t)}, \tag{5.3}$$

where $n^K_{od}(t)$ (unit: veh) and $m^K_{od}(t)$ (unit: veh·km) are the accumulation state and the total remaining distance to be traveled by vehicles in state K in region o destined to d, respectively, $I^K_{od}(t)$ (unit: req/h) indicates the summary of inflow rates, including the start of new trips ($\lambda^K_{od}(t)$) and the transfer of vehicles reaching a new region while performing their activities ($O^K_{od}(t)$), $\alpha_{od} \geq 0$ is a parameter expressing sensitivity of the transfer flow to variations in the remaining distance to be traveled, $L^K_{od}(t)$ (unit: km) is the average regional trip length, $L^*_{K_{od}}$ (unit: km) is the average remaining trip length in steady-state (with $L^* = (L^2 + \sigma^2)/2L$, where σ is the trip length standard deviation), while $v_o(t)$ (unit: km/h) indicates the traveling speed.

To apply the MPC control inputs in a two-region, Equations [5.4a]–[5.4c] detail the dynamics of vacant ride-hailing vehicles with the applied control parameter $r_{od}(t)$.

$$\dot{n}^I_{oo}(t) = O^B_{oo}(t) + O^V_{ho}(t) - \sum_{d \in \mathscr{R}} \bar{\lambda}^B_{od}(t) - r_{oh}(t) \qquad h \neq o \tag{5.4a}$$

$$\dot{n}^I_{od}(t) = r_{od}(t) - O^B_{od}(t) \qquad d \neq o \tag{5.4b}$$

$$\dot{m}^I_{od}(t) = r_{od}(t) L^I_{od}(t) - n^I_{od}(t) v_o(t) \qquad d \neq o \tag{5.4c}$$

where $\bar{\lambda}^B_{oh}(t)$ is the passenger entrance rate as described in Beojone and Geroliminis, 2023a.

The dynamical equations given in Equations [5.1] and [5.2] can be discretized in time with a sampling time T_u (unit: h), for enabling formulation of an associated finite-dimensional nonlinear optimization problem. Rewriting them in a compact form, we obtain the following vector nonlinear equation:

$$x(k+1) = F(x(k), q(k), u(k)), \tag{5.5}$$

where $k \in \mathbb{N}_0$ is the time step of sampled real time (i.e., $t(k) = T_u \cdot k$, $k = 0,1,2,\ldots$), $x(k) \in \mathbb{R}^{n_x}$ (state) is a vector containing all M-model state variables (i.e., accumulation states $n^P_{od}(k)$, $n^B_{od}(k)$, and $n^V_{od}(k)$, total remaining distance to be traveled states $m^P_{od}(k)$, $m^B_{od}(k)$ and $m^V_{od}(k)$). $q(k) \in \mathbb{R}^{n_q}$ (measurable disturbance) is a vector containing all exogenous demands (i.e., $q^P_{od}(k)$ and $q^B_{od}(k)$), whereas $u(k) \in \mathbb{R}^{n_u}$ (control input) is a vector containing all relocation control input terms $r_{od}(k)$.

Based on the dynamical model Equation [5.5], we can formulate the problem of finding the optimal relocation control input $r_{od}(k)$ values that minimize the total number of canceled trip requests, as the following discrete-time economic nonlinear MPC problem:

$$\min_{u_\kappa} \quad \sum_{\kappa=1}^{N} \sum_{o\in\mathscr{R}} \sum_{d\in\mathscr{R}} \lambda^B_{od}(k+\kappa) \exp(\gamma_0 (n^V_{od,\kappa})^{\gamma_1} v^{\gamma_2}_{o,\kappa} w^{\gamma_3}) T_u \tag{5.6}$$

$$\text{s.t.:} \quad x_0 = x(k) \tag{5.7}$$

$$\text{for } \kappa = 0,\ldots,N-1: \tag{5.8}$$

$$x_{\kappa+1} = F(x_\kappa, q(k+\kappa), u_\kappa) \tag{5.9}$$

$$0 \le Tr_{od,\kappa} \le n^V_{od,\kappa} \text{ for } o,d \in \mathscr{R}, \tag{5.10}$$

where κ is the MPC time interval index (i.e., discrete-time clock internal to the MPC), k is the current discrete time step, N is the prediction horizon, while x_κ and u_κ are the state and control input vectors internal to the MPC (i.e., predicted states and controls), respectively. To compute the loss probability, $\lambda^B_{od}(k)$ is an exogenous demand inputs for busy ride-hailing vehicles; the remaining $\gamma_i, i \in \{0,1,2,3\}$ are parameters expressing the sensitivity of the matching algorithm to the number of vacant vehicles ($n^V_{od,\kappa}$), traveling speeds ($v_{o,\kappa}$) and passenger waiting time tolerance (w). A detailed estimation of parameters in Problem [5.6] will be provided in the full paper.

5.2.2 Middle-layer: Selecting and dispatching vehicles

Once the upper-layer provides the number of vehicles transferring between regions, a further step is to select which vehicles should move to other desired regions (i.e., repositioning vehicles). Simultaneously, vacant vehicles staying in the current region (i.e., idle vehicles) are operated to maintain a good spatial configuration to uphold service quality. In summary, for each region R, the middle-layer is responsible for bridging the results from the upper- and lower-layers, such that inter- and intra-regional instructions are optimized.

Hence, besides the information acquired from the upper-layer, the middle-layer needs an estimate of the coverage instructions in the lower-layer described in Section 5.2.3. We propose employing, for these estimates, the optimal configuration of intra-regional positions (centroids) used in the lower-layer. Note that this coverage must be evaluated only for vehicles

staying in the current region. Therefore, we sample random u_{RR} positions in region R (where $u_{Rr} = r_{Rr}(k) \cdot T_u \forall r \in \mathscr{R}$ is the conversion of the outputs from the upper-layer) and compute the centroids that maximize the coverage as described in Section 5.2.3.

Then, given the computed intra-regional vehicle positions and the number of repositioning vehicles to each region, the middle-layer is able to perform the assignment of vehicles by solving the optimization problem in Equation [5.11], complying with the upper-layer decision (i.e., the number of idle and reposition vehicles $u_{Rr} = r_{Rr}(k) \cdot T_u$). Specifically, for objective D_R, the first term considers the reposition distance between regions, while the second term takes the intra-regional traveling distance into account:

$$\min_{\psi_{ir},\omega_{il}} D_R = \sum_{i\in I}\sum_{r\in\mathscr{R}} \psi_{ir} d_{ir}^{\text{out}} + \sum_{i\in I}\sum_{l=1}^{u_{RR}} \omega_{il} d_{il}^{\text{in}} \tag{5.11a}$$

$$\text{s.t.:} \quad \sum_{r\in\mathscr{R}} \psi_{ir} + \sum_{l=1}^{u_{RR}} \omega_{il} = 1 \qquad \forall i \in I \tag{5.11b}$$

$$\sum_{i\in I} \psi_{ir} = u_{Rr} \qquad \forall r \in \mathscr{R}\backslash\{R\} \tag{5.11c}$$

$$\sum_{i\in I} \omega_{il} = 1 \qquad \forall l \in \{1,\ldots,u_{RR}\} \tag{5.11d}$$

$$\sum_{i\in I}\sum_{l=1}^{u_{RR}} \omega_{il} = u_{RR} \tag{5.11e}$$

$$\psi_{ir} \in \mathbb{B} \qquad \forall i \in I \text{ and } \forall r \in \mathscr{R} \tag{5.11f}$$

$$\omega_{il} \in \mathbb{B} \qquad \forall i \in I \text{ and } \forall l \in \{1,\ldots,u_{RR}\} \tag{5.11g}$$

where ψ_{ir} and ω_{il} are binary decision variables expressing whether a vehicle i is assigned to region r or to position l; d_{ir}^{out} and d_{il}^{in} are the traveling distances for a vehicle i to reach region r or the position l in the current region, respectively; I is the set of vacant vehicles; u_{Rr} is the control action obtained in the upper-layer, indicating how many vehicles should move from region R to region r, while u_{RR} is the number of idle vehicles that should stay in the current region.

Note that the constraint in Equation [5.11b] limits a vehicle either to stay in the current region or move to another region. Equations [5.11c] and [5.11e] ensure compliance with the upper-layer decision. This problem was inspired by classic assignment problems, such as those in Pentico, 2007.

Assignment problems are fundamental combinatorial optimization problems. Note that the number of tasks (destinations inside and outside the evaluated region) is $|\mathscr{R}| - 1 + u_{RR}$, while number of agents (idle vehicles in the evaluated region) is $\sum_{r\in\mathscr{R}} u_{Rr}$. One could see the previous as an unbalanced assignment problem for having different number of tasks and agents. However, differently from the classical assignment problem, the proposed problem assigns

multiple (u_{Rr}) vehicles to inter-regional tasks. Therefore, if one repeats u_{Rr} times each of the columns of the cost matrix associated with inter-regional assignments, the problem becomes balanced (number of agents and tasks become equal). Finally, as typical to assignment problems, the solution can be obtained using the Hungarian method (Kuhn, 2005) in strong polynomial times (Munkres, 1957), without the use of dummy agents nor tasks.

A possible solution of the problem in a two-region setting can be illustrated in Figure 5.2. It presents a situation where some vehicles from a Region 1 are assigned new positions in their current region and the remaining ones are sent to Region 2 to the lower-layer controller perform the local repositioning instructions. We must highlight that vehicles are not sent to the center of Region 2, they are, in fact, dispatched only to the border between Regions 1 and 2 by the shortest path. Therefore, the lower-layer will be able to instruct these vehicles as quickly as possible. Moreover, note that the vehicles in the current region are chosen such that the cost (total traveled distance) to reach the desired covered is minimized, enabling quick intra-regional response to unbalanced demand coverage, as well.

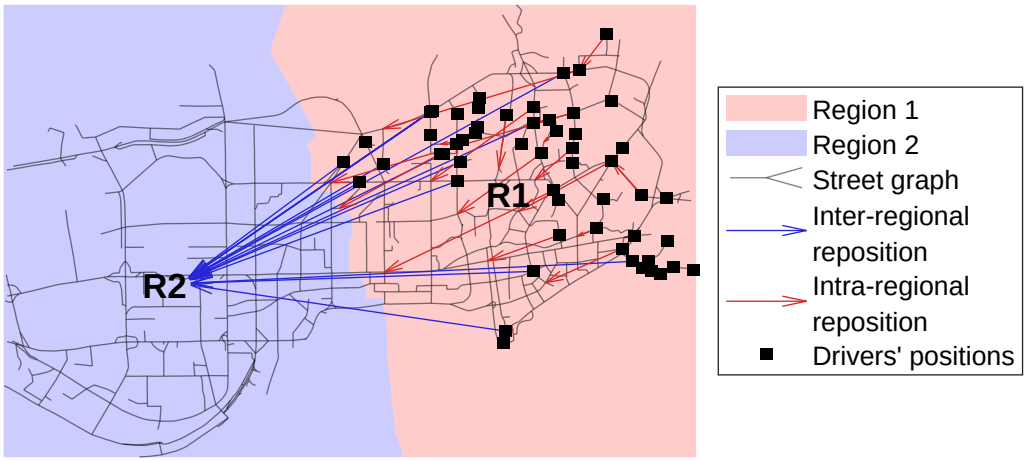


Figure 5.2: Illustration of the solution for the Middle-layer in the form of an assignment problem. Red arrows indicate the exact position a vehicle is assigned in intra-regional movements. Blue arrows only indicate the region of destination.

5.2.3 Lower-layer: Coverage control method

In the lower-layer, the coverage control algorithm is operated for the vehicles that are commanded to stay in the current region (i.e, idle vehicles). The coverage controller steers these vehicles towards an optimal spatial configuration (indirectly, towards maximizing availability for service) by operating at a fast time scale and with detailed position guidance.

For each region, we formulate this task as a coverage control problem for the coordination and deployment of multiple mobile agents on the city network (Zhu et al., 2022). Such coordination provides benefits to the system by dynamically allocating the vehicles according to the different

demand densities of various city districts.

The city map can be presented as an undirected graph $G = (Q, E)$, where Q is the set of nodes representing the intersections and E is the set of road links. If origin-destination pairs for trips are recorded in historical taxi data, we can compute the probability that a request starts at a node as $\phi(q)$. With a slight abuse of notation, q in this section denotes a node on the graph (with $\sum_{q \in Q} \phi(q) = 1$).

The city area is clustered into R_n regions, with the set of all nodes in region R denoted as Q^R, where $Q = Q^1 \cup Q^2 \cup \cdots \cup Q^{R_n}$. For idle vehicle i in region R, whose current position is p_i, the Voronoi tessellation can be defined as follows (Erwig, 2000):

$$V_i(p_i) = \{q \in Q^R : d(p_i, q) \leq d(p_j, q)), \forall\, i \neq j\}, \tag{5.12}$$

where $d(p_i, p_j)$ stands for the shortest distance between node p_i and node p_j on graph, computed by the Floyd-Warshall algorithm and $P = \{p_1, p_2, \ldots p_{n_{idle}}\}$, where n_{idle} is the number of current idle vehicles in region R.

The coverage objective function can be formulated as:

$$H(P, V) = \sum_{i=1}^{n_{idle}} \sum_{q \in V_i} d(p_i, q)^2 \phi(q). \tag{5.13}$$

According to Durham et al., 2012, the optimal position configuration of all vehicles is attained when each vehicle is at the centroid of its respective Voronoi cell. The centroid of a graph Voronoi cell can be computed as an integer optimization problem as

$$C(V_i) = \arg\min_{q} \sum_{q \in V_i} d(p_i, q)^2 \phi(q). \tag{5.14}$$

At the beginning of each fast-loop time period T_l, the idle vehicle i is commanded to move towards its current Voronoi centroid $C(V_i)$ after solving Equation 5.14. Note that the number and position of vehicles will change (new assignments or service completion), which requires us to re-compute the Voronoi diagram completely at each iteration (which is done in low computational times using parallelization). As it only requires local information for each vehicle to calculate the Voronoi tessellation, this control algorithm is able to provide each vehicle with an intersection/node-level rebalancing command in a distributed manner (i.e., without requiring a central planner to coordinate the movements of all vehicles).

5.3 Computational Results

To evaluate the proposed strategy, we use an agent-based simulator which is adapted from (Beojone & Geroliminis, 2021b). With sufficiently long and rich historical data, future requests can be predicted accurately. So far, we assume that we have perfect knowledge of the demand profile. The road network for the Futian and Luohu districts of Shenzhen, China, forms the background for the study. The considered network consists of 1'858 intersections connected by 2'013 road segments. In total, the origin-destination demand data contains around 200'000 requests collected from taxi operations using GPS coordinates (Ji et al., 2014). Using a graph-based k-means algorithm the intersections are clustered into two urban regions.

In total, we tested 5 demand scenarios with 250 vehicles. We consider a three-hour simulation, where, in the base demand scenario, 600 requests per hour are issued in the first and third hours and 1'200 requests per hour are issued in the second hour. In each of the subsequent scenarios, the base demand is increased by 25%, reaching a 100% increase in the last scenario. Larger fleet sizes and multiple (i.e., more than two) region cases will be studied in the full paper. The upper-layer is activated at a slow time scale, specifically, $T_u = 10$ min, while the coverage control in the lower-layer works at a fast scale with $T_l = 10$s.

Note that the middle-layer can only be applied when the upper-layer is active, therefore scenarios with active upper-layer (MPC) also include the middle-layer. In summary, we tested four scenarios: (i) full framework (called 'MPC+CC'), (ii) only the lower-layer coverage control (called 'CC only'), (iii) only the MPC and middle-layer (called 'MPC only'), and (iv) no active controls (called 'Do nothing').

Each control layer individually managed to improve the average waiting times and the acceptance rates. Especially, the proposed hierarchical control framework, combining all layers, yields the best performances over all scenarios with the lowest waiting times and highest acceptance rate. Figure 5.3 illustrates the preliminary results when applying different repositioning strategies. For instance, in the original demand scenario, average waiting times were 25% lower in comparison with a 'Do-nothing' scenario and accepted approximately 14% more requests. For the upper-layer, the MPC controller is designed for minimizing cancellation, which consequently inspires to answer more requests; for the lower-layer, the coverage control scheme steers the idle vehicles to move toward the high-demand areas, where the passengers can benefit from easily finding an available vehicle around.

Figure 5.4 illustrates the number of vacant vehicles in each region governed by the hierarchical control strategy (all 3 layers developed in Sections 5.2.1–5.2.3) *vs* do-nothing strategy, when increasing 50% demand. It can be seen that during the peak period, there are few vacant vehicles in both regions with the hierarchical control framework, which reveals that the proposed method can operate the fleet efficiently and most of them are actively serving passengers. Once the reposition vehicles arrive in their destinated regions, they become 'idle' again. Therefore, the number of reposition vehicles decreases fast to 0 thanks to the middle-layer controller tending to select the vehicles that are not far away from the destinated

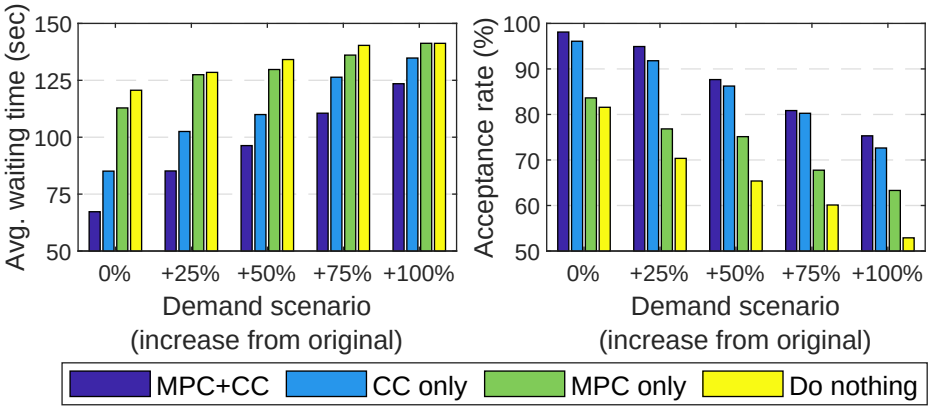


Figure 5.3: Average waiting times and acceptance rate in each demand scenario.

region, thus the reposition ones will reach the desired regions soon. However, many vehicles are not under use with 'Do-nothing' strategy. And more and more vacant vehicles accumulate in low-demand areas in region 1, leading to a greater amount of request cancellations.

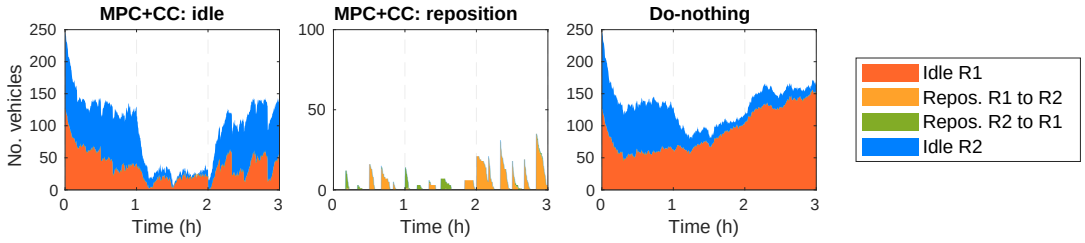


Figure 5.4: Evolution of the number of vacant drivers in each region under the proposed 'MPC+CC' strategy and a 'Do-nothing' strategy (scenario: 50% demand increase).

5.4 Summary

In this chapter, we investigated a hierarchical control framework for repositioning empty vehicles. For the upper level, we introduced an MFD-based model to describe the dynamics of both the taxi and private vehicles and an MPC controller is designed to instruct the transferring repositioning of vehicles between regions; for the lower layer, the fleet location optimization problem was solved as a coverage control problem, which can be carried out by each vehicle to generate its own position guidance. To bridge between the upper and lower layers, an assignment problem minimized the rebalancing cost of complying with the objectives of both layers. The effectiveness of the proposed method is verified by an agent-based simulator modelling the real road network of Shenzhen, China. The proposed hierarchical framework yields its advantages by serving more passengers with less waiting time.

6 Conclusions

This thesis improved our understanding about ride-sourcing services, its interactions in the urban area, with respect both to drivers and passengers, while keeping on perspective the surrounding sustainability dimensions. The focus grew from the understanding of effect of expanding ride-sourcing in traffic in chapter 2, passing through its macroscopic modeling in chapter 3, while chapter 4 evaluated the potential repositioning response when using revenue forecasting and 5 proposed a hierarchical control strategy for relocating idle ride-sourcing vehicles. This final chapter presents a summary of the conducted research, highlighting the main findings and contributions while proposing promising areas of related future research.

6.1 Main findings and discussions

In Chapter 2, we investigated the effect of expanding fleet sizes for TNCs, passengers with different willingness to share, and operational strategies over congestion conditions. We highlight that, by omitting the dynamics of congestion in rides-sourcing studies to focus on matching strategies or rebalancing vehicles in static environments, different conclusions with possibly unrealistic performance measures are obtained.

Results show that sharing (allowing ridesplitting with a large pool of passengers) by itself is not capable of decreasing the system's VKT if there is no control over the fleet (its size and operation). To reduce emissions (by reducing VKT), TNCs should change their *modus operandi*; in a way to avoid that their fleet cruises without an assigned passenger. On the other hand, sharing decreases the number of vehicles needed to maximize coverage and minimize service times. Furthermore, in case it is not possible to avoid TNCs' fleets cruising for passengers, increases in the willingness to share can minimize both waiting times and service times. For adequately sized fleet sizes, the adoption of sharing is related to higher revenues (for the system and the driver). Therefore, so that ride-sourcing becomes a sustainable service, it must change its operations to remove vehicles without passengers from the streets, and passengers must become more receptive to ridesplitting at the same time. Furthermore, our findings show that TNCs' operations with large fleet sizes can lose attractiveness to public transportation

as a result of high traffic congestion, and drivers can have attractive revenues even in such situations. Nevertheless, we considered no interactions between these modes, and, even with dedicated bus lanes, there would be interactions and problems on intersections. It is outside the scope of this paper to combine surge pricing policies and market equilibrium, but this can be a future direction. Therefore, we do not cope with the detailed modeling of decisions on the mode choice for travelers.

In chapter 3, we proposed a dynamic model capable of representing ride-sourcing services and private vehicles macroscopically in an urban network separated in a multi-region setting. We supplied the processes for estimating the required parameters and computation of errors. It depicted mass conservation equations for both ride-sourcing and private vehicles (backgrground traffic).

We evaluated a multi-region setting and its sensitivity to passengers' willingness to share, their waiting time tolerance, and ride-sourcing drivers' fleet size. The model directly captured the effects of wild-goose-chase (Castillo et al., 2018), confirming that limiting matching radius – in this case, represented by decreasing waiting time tolerance – hinders its effects, decreasing the overall waiting time and number of lost requests, even in transient scenarios (Xu, Yin, & Ye, 2020; H. Yang et al., 2020). The multi-regional setting could further limit pick-up distances by limiting the matching radii, which could be upper bounded by well-known methods based on the region area and street network topology (see Larson and Odoni, 1981, Chapter 3, for instance).

Assuming an FCFS matching scheme, the model showed the fleet size and willingness to share as crucial to match travelers into a single vehicle. Forming a pool of passengers and having fewer drivers to serve them forces the operator to bring travelers together. Furthermore, observing the regional dynamics further allowed identifying the influence of demand patterns on local driver availability and instantaneous traffic conditions represented primarily by the background traffic of private vehicles.

The final evaluation tested the model quality to different prediction horizons, which presented increasing errors to their lengths but remained below 10% in all cases. In the next step, we compared the proposed model to different benchmarks, and the computed errors were only a fraction of theirs. Like the exceptional error measurements, the actual values closely followed the plant at all times in a test built to mimic the rolling time-horizon structure of an MPC controller. Moreover, the proposed model had lower error measurements than the benchmarks in all states. Representing a unique state encompassing pick-up and delivery activities for each ride-sourcing service decreased the sensitivity of the proposed model to measurement noises. Precise predictions for ride-sourcing in dynamically congested areas present the next step towards better traffic control and service operations management.

We performed additional tests on the model, evaluating its sensitivity to noises in the inputs of private vehicles. However, for scenarios with an average noise of 15% or larger, total errors decreased at longer prediction horizons. Noises were limited to measurements of the

starting number of private vehicles, assuming that, with current technologies, one can have all information needed from ride-sourcing vehicles. Even in a scenario with advanced technology employed in collecting precise traffic data, stability and robustness must be regarded as vital characteristics of models when constructing real-life solutions. Based on our findings, besides the mobility benefits of ridesplitting, it also increases model stability, decreasing the sensitivity to noise. In light of such findings, as models' complexity can increase, careful and systematic analysis of prediction errors and sensitivity to noises in particular measurements pose an unavoidable step for model evaluations.

Chapter 3 is among the first attempts to present and evaluate such a model with ridesplitting (shared rides) in the literature. The success of a dynamic model with ride-sourcing provides scholars, practitioners, and authorities a tool for measuring the interactions such services over traffic in a simulated and fast environment and proper for studies improving shared rides in a congested urban area. Further developments on repositioning strategies ride-sourcing services with ridesplitting options and traffic congestion can profit from dynamic models such as the one proposed in this chapter. Notably, the proposed model could support high-level repositioning decisions in hierarchical problems like the one seen in Yildirimoglu et al., 2018. Furthermore, problems in strategic market/regulatory responses to ride-sourcing services in transient situations can employ the proposed model.

Therefore, in chapter 4 we proposed a relocation strategy for ride-sourcing drivers by providing them with an estimate of their earnings. Therefore, we do not assume drivers unrestricted compliance to the provided guidance and, thus, they are free to make the decision that they expect will maximize their earnings. The first step in the proposed approach uses simulation-based optimization to identify the locations that are expected to maximize the chances of a driver getting a match in the forecast horizon. Then, a MDCTMC model is developed to capture the activities a driver will perform depending on his/her decision, which is later translated into an estimate of the driver's earnings. We showed that the proposed approach is likely to retain drivers confidence by improving their earnings compared to other drivers if the operator selects only a fraction of active drivers to provide guidance.

Besides improving earnings, we show that the proposed approach manages to decrease the number of unserved requests in the system compared to several state-of-art benchmarks. It increased vehicle occupancy, and decreased the deadheading, while enhancing driver's compliance to the most profitable decisions.

These findings are consistent with the previous works of Ramezani and Nourinejad, 2018, Nourinejad and Ramezani, 2020 and Xu, Yin, Chao, et al., 2020 that were able to find improvements to different objectives in taxi services by repositioning drivers dynamically. Moreover, they provide a path for testing the impacts of different regulatory schemes in such systems. It enables improved services from the perspective of the service operator at the same time that provides tools for regulators to identify areas with poor coverage, which could receive governmental incentives to provide the local population with improved mobility and accessi-

bility options. For instance, identifying the best places to have a subsidized driver providing mobility to population with lower income or limited mobility to access other public transport options.

Among the identified limitations, we must highlight that by focusing on revenues drivers can find individually better outcomes but not necessarily improving service quality to its full potential. In other words, there is a Price-of-Anarchy associated to the strategy, which can be identified in this study, but such an evaluation would require more in-depth observation and development to mitigate potential negative externalities. Furthermore, in one hand using simulation-based optimization brought flexibility to the definition of the ideal repositioning destinations and path. In the other hand, it can become a bottleneck in applications at larger and denser urban areas. In this direction, simplifying the simulation used, or offering other optimization strategies (e.g. surrogate optimization, as seen in Chen et al., 2019, for instance) could improve its transferability to any scenario. Furthermore, instead of using the shortest path, one could expand it to the k-shortest paths, further increasing the options for higher chances of matching.

The provided guidance could further benefit drivers and unserved passengers, if it comes paired with other mechanisms to foster movements to poorly covered areas. It could include lower commissions in these areas, or other price changes to make it more attractive to drivers. Other research directions include developing optimal control to reposition without the decision-making process by drivers, which would be more realistic in cases with autonomous vehicles but it would also serve as upper (lower) bound for performance measurements and evaluation. In this case, it is interesting to evaluate how the different objectives of the controller can impact different measurements of revenues, abandonments, and changes in traveler and driver behaviors.

In chapter 5, we presented a hierarchical control framework capable of repositioning vacant ride-hailing vehicles integrating model predictive control and coverage control in an urban traffic setting. The approach involves optimizing vehicle positions using near-future forecasts of the service and demand conditions, presenting a proactive strategy for dynamically deploying the fleet in advantageous spatial configurations.

Each layer individually managed to improve waiting times and acceptance rates, increasing the number of served requests and improving service quality. The combined effect of each layer further improves the performance in all tested scenarios. It is worth mentioning that the combined effect in the average waiting time is not linear. In other words, the average waiting time decreased more when all layers are combined than the sum of the decrease of each individual layer, which illustrates a positive feedback in combining different strategies.

The proposed framework also maximized the use of the available fleet. For instance, during the peak hour it managed to keep almost no idle vehicle in any region while maintaining the good level of service mentioned before. Moreover, the results showed that drivers would fulfill their repositioning instructions fast, such that almost no vehicle was on inter-regional

movement when the upper- and middle- layers were called. That illustrates that the middle-layer managed to identify responsive assignments, even when a considerable number of drivers was selected to perform inter-regional movements.

Among the limitations we identified, we must highlight that the upper-layer might overreact to imbalanced scenarios, e.g., sending more repositioning vehicles than is truly needed, alternating regional flows in a few successive steps. It may be caused by the passenger-driver matching in the prediction model which lacks integration to the lower layer and might underestimate the covered demand. Therefore, further study should calibrate the modeled matching process depending on the amount of movement to complete the coverage decisions. Regarding the middle layer, we designed the objective function with equal weights on the cost by repositioning some vehicles to another region and by moving the rest of them to optimal configuration within the current region, and further study is also needed to balance the trade-offs between these two.

6.2 Future research directions

Based both on the developments and findings of the research included in this thesis and the emerging changes in the urban mobility solutions, interesting future research directions emerge related to the research field explored.

Chapter 2 identified a potentially harmful scenario for urban mobility as a consequence of ride-sourcing activities, an increased level of congestion due to a large number of circulating vehicles for passengers. At the same time, it tested idealized solutions, such as providing a parking option for drivers and expanding the usage of ridesplitting. The later chapters in this thesis also expanded the management of the service fleets improving the service but not accounting for mitigating the negative externalities of having oversized service fleets cruising on the streets leading to deteriorated urban traffic conditions. Additionally, the service operators still struggle to encourage a broader adoption of ridesplitting by the passengers. Identifying methods that increase passengers' adoption of shared modes (such as ridesplitting) will open fields of research on practical problems such as improved matching algorithms, dial-a-ride problems, so on. It is especially true in a multi-modal scenario where authorities try to promote public transit as a solution to congestion problems.

In this direction, one could explore methods to attract passengers to ridesplitting, especially during the most congested hours of the day. Ride-sourcing services may experience dynamic taxation on empty vehicles operating inside a congested area. On the other hand, they may experience no taxation in areas with low congestion and with an unsatisfactory public transit offer for users. Another option would be to understand whether occupation-based pricing could make ridesplitting a viable alternative during peak hours and improve social welfare. Given the complexity of such a scenario, results may rely on simulation approaches, instead of analytical solutions. Any sort of regulations in such service would primarily affect passengers' demand by changing their costs, and affect drivers' supply changing their revenues.

Passenger choices can focus on minimizing their traveling costs, while drivers can join based on an opportunity cost (Vignon et al., 2021). In aggregated traffic dynamics, in the form of a Speed-MFD and single region settings, traffic assignment is secondary to the problem. Fixed taxation could obtain unfavorable outcomes for the society, especially during peak hours hindering the service and the ridesplitting option simultaneously. A non-linear pricing strategy (Lawphongpanich & Yin, 2012), in the other hand, could overcome this issue by leveraging the charges in the correct moments to maximize welfare by fostering ridesplitting.

In a more operational oriented approach, one could work promoting ridesplitting dividing passengers in groups: (i) Selfish users that do not share their information and their rides; (ii) Partially cooperative users that share their information or their rides (but not both); and (iii) Cooperative users that share their information and their rides. The operator gives a discount based on the cooperation of the user and commits on delivering him/her on time. Passengers may share their information in advance, allowing scheduling from the operator. The details of the implementation of this scheduling service remain vague. However, we acknowledge that lost freedom should be associated with a cost. We call this cost as "cost of cooperation," and it may consider the sum of an expense associated with scheduling, and another one that is particular to the technology employed (e.g., the website used for time slot reservation). We assume that individuals will get a distributed cost of cooperation, but it will remain constant for one individual over time. These costs will act as the selecting criterion: at equilibrium, users only choose to cooperate, if, by doing so, they can reduce their expense associated with the travel time and schedule penalty by more than their cost of cooperation. Note that a similar concept is used to distinguish planning users from not planning users in previous works related to public transit (Fosgerau, 2009; Lamotte et al., 2017; Tisato, 1992). In other words, one may consider the proposed service as an additional alternative between public transit and private vehicles.

In a different direction, urban mobility continues to evolve and new solutions emerge constantly. For instance, unsatisfied with the conditions to work with Uber, drivers organized themselves in a few cities in Brazil and, in a partnership with the local authorities, launched their ride-sourcing service to compete with Uber, recently (Gama, 2022) In other cases, the local authorities themselves developed and launched their ride-sourcing platform as a "market regulating element" (in Portuguese, "*elemento regulador do mercado*") (Band Jornalismo, 2023). In the first case, the company can operate as a cooperative of drivers, where the profit is shared among the cooperates. While, in the second case, the formed organization can be operated such that social welfare is maximized. In both cases, their strategy to attract drivers is to have lower commissions for rides. To attract passengers, they mention that there is no surge pricing applied in peak hours. Since these services have an operator related to the local authorities, connecting it with existing public transit modes should become easier in the context of Mobility as a Service. Such an unified solution would also meet the idea of smart cities with connected transportation modes.

As an emerging service, it is still unclear whether this operation will be able to compete, for

passengers and drivers, in the long term. For instance, in their first day of operation, there were more then 30'000 registered passengers but they were lacking drivers because of a citywide surge in demand due to a labor strike in metro services (Ker, 2023). Therefore, to achieve (if possible at all) the benefits presented in the previous paragraph, one must understand where such a service will stand in this competition. Furthermore, one must be able to identify in which cases this service can become prevalent, if it will create a monopoly, or if it will not manage to survive in the long-term because the competitors are well-established with a large pool of drivers and passengers. With a game-theoretical approach, analytical solutions could provide the existence and uniqueness of equilibrium, if any at all. The literature on conventional ride-sourcing services has advanced significantly in the recent years, with studies on ride-sourcing services in monopoly and duopoly conditions (Ke et al., 2021; Zha et al., 2016; K. Zhang & Nie, 2021). However, given the different nature of this new service, it is unclear how its organizational structure would change the objectives and how revenues and costs are perceived. For instance, since it parts from drivers organizing themselves, profit could be proportionally shared among them (similarly to a cooperative structure), while the operation has a fixed cost related to the maintenance of infrastructure provided by local authorities.

It is also unclear how this organizational structures could boost or undermine its potential in achieving competitive service and integration with other modes. In such a case, qualitative research is needed to classify this organization and identify possible strengths, weaknesses, opportunities and threats (in reference to the management literature in Weihrich, 1982). In this direction, a traditional case study could combine data from structured interviews and questionnaires, internal documents evaluation, and direct observations. The idea of having different sources of data is to avoid biases in each individual data set and identify overlapping information for consistency (Yin, 2018).

A MFD modeling basics

An MFD provides a well-defined empirical relationship between the number of vehicles in an area (accumulation) and its average speed or the total traveled distance per unit of time (production). In summary, it can be expressed as $n \mapsto v(n)$ or $n \mapsto P(n)$, where n, represents the accumulation, while $v(n)$ and $P(n)$ represent the average space-mean network speed and production, respectively. Note that one can obtain the $P(n)$ relationship by the product $P(n) = n \cdot v(n)$. These representations can be called "speed-MFD" and "production-MFD," respectively (Lamotte & Geroliminis, 2018).

On average, all drivers would complete their trips or exit the hypothetical area after traveling a certain distance L. Therefore, assuming that inputs to that area change slower than the relaxation time (time to the travel across the region), the exit function could be written as $O(n) = P(n)/L$ (Daganzo, 2007). Note that an exiting function may represent trips that end (trip-completion) or leave (transfer) the hypothetical area.

Under an input function $\lambda(t)$ and an initial condition $n(0)$, the dynamics of the number of drivers can be described in Equation [A.1], also called the mass-conservation equation (Daganzo, 2007; Mariotte et al., 2017). Given the derivation of the exit function, this formulation is usually called the accumulation-based model.

$$\mathrm{d}n(t)/\mathrm{d}t = \dot{n}(t) = \lambda(t) - O(n(t)) \tag{A.1}$$

A.1 Trip-based model

Consider a driver m with a trip length of L_m entering the system at time t_m; then this driver exits the system after travelling τ_m time units as computed in Equation [A.2].

$$L_m = \int_{t_m}^{t_m+\tau_m} v(n(s))\,\mathrm{d}s \tag{A.2}$$

Note that the speed $v(n(s))$ results from a speed-MFD. In other words, the trip-based model considers the traveled distances explicitly, not requiring a particular $L_m = L$ average trip length, resulting in more accurate dynamics for transient situations than the accumulation-based model (Paipuri & Leclercq, 2020). However, it becomes difficult to solve analytically, as indicated when introduced in Arnott, 2013. For this reason, Mariotte et al., 2017 proposed an event-based approach to obtain numerical solutions. If trip lengths are exponentially distributed, trip-based and accumulation-based models are identical. Studies mainly propose this model to investigate departure time choice problems at the city scale (Arnott, 2013; Batista & Leclercq, 2019; Daganzo & Lehe, 2015; Fosgerau, 2015; W.-L. Jin, 2020; Lamotte & Geroliminis, 2018; Leclercq & Paipuri, 2019; W. Vickrey, 2020).

A.2 Intermediate approach: M-Model

First introduced in Murashkin, 2021, the M-model tries to overcome the limitations of the accumulation-based model by summarizing the past events into the total remaining distance M and using this information to update the exit function. For comparison, trip-based models keep track of individual remaining distances, while accumulation-based models keep no record of it. Such a model offers valuable intuition and represents an attractive trade-off for control applications. Sirmatel et al., 2021 provided a multi-region formulation of the M-model and integrated it successfully in a perimeter control framework.

One can derive the dynamics for computing the total remaining distance based on the dynamics of the accumulation-based model. Assuming that the average trip length can represent the added remaining distance for each entering driver; then, one can obtain Equation [A.3].

$$\mathrm{d}M/\mathrm{d}t = \dot{M}(t) = \dot{n}L = \lambda(t)L - P(n(t)) = \lambda(t)L - n(t)\,v(n(t)) \tag{A.3}$$

To update the exit function, Lamotte et al., 2018 proposed using a correction factor in the form of Equation [A.4].

$$O(t) = \frac{n(t)\,v(t)}{L}\left(1 + \alpha\left(\frac{M(t)}{n(t)L^*} - 1\right)\right) \tag{A.4}$$

where L^* is the average remaining distance at steady-state (computed as $L^* = (L^2 + \sigma^2)(2L)^{-1}$,

where σ is the standard deviation of trip lengths), and α is a constant parameter related to the distribution of trip lengths (see Lamotte et al., 2018 for more details on setting this constant).

Therefore, the M-model uses Equations [A.1] and [A.3] (replacing the exit function with the result of Equation [A.4]) to keep track of the number of vehicles in the area and their total remaining distance. For more details about the properties of the M-model, the reader could refer to Murashkin, 2021.

B Monte Carlo Simulation

A Monte Carlo simulation evaluated the influence on the non-acceptance (loss) of incoming requests after checking the feasibility constraints (waiting time and/or detour) of factors such as the available fleet size $n_{od}^{av}(t)$, the regional average speed $v_o(t)$, the waiting time tolerance ω, and the Idle-Busy drivers' ratio $\rho^s = n^I(t)/n_{od}^{av}(t)$. We simulated loss probabilities for the variable tuple $(n_{od}^{av}(t), v_o(t), \omega, \rho^s)$. Where we tested all possible combinations of the following values for each variable:

- $v_o \in \{5, 10, 15, 20, 25, 30, 35, 40\}$ (unit: km/h);
- $n_{od}^{av} \in \{10, 30, 60, 110, 150, 210, 270\}$ (unit: #veh);
- $\omega \in \{2, 5, 8, 11, 14, 17, 20\}$ (unit: min); and
- $\rho^s \in \{0.1, 0.3, 0.5, 0.7, 0.9, 1\}$ (unitless).

Note that $\rho^s = 1$ will make the problem indifferent in terms of service since all drivers are idle.

The simulation considers a street network divided into regions, depending on the desired case. Since drivers and passengers must be in the same region to perform an assignment, each region had an independent Monte Carlo simulation. Under the assumption of an FCFS discipline without batching, it always yields a higher supply of drivers than demand for passengers.[I] Moreover, separate simulations evaluated ride-hailing and ridesplitting services. Algorithm 1 shows a pseudo-code that illustrates the Monte Carlo simulation. We sample and test numerous potential passengers ($pasSamp = 500$) individually, considering all vehicles in the sampled fleet. Furthermore, to ensure that there is no bias in the vehicle positioning, we sample several combinations of positions for drivers ($vehSamp = 20$) for each set of parameters. Four properties describe each of the sampled drivers:

- Current position ($.curr$): Node in the street network;

[I] See Footnote IV.

- Origin position ($.orig$): Node in the street network – empty for idle drivers;
- Destination position ($.dest$): Node in the street network – empty for idle drivers;
- Busy flag (boolean): 'false' for idle drivers, 'true' for busy ones;

The evaluation of the waiting time and detour tolerances are described below in Algorithm 2 to detail the matching constraints evaluated in the function in Algorithm 1. In this process, besides the main parameters, we use the shortest distance between two points in the street network $allDist(\cdot,\cdot)$, previously computed using a Floyd-Warshall algorithm, and a maximum detour tolerance $detour$. Note that the evaluated constraints for ride-hailing and ridesplitting are the same as in (Beojone & Geroliminis, 2021b).

Data: $allF$, // Set of all tested fleet sizes
$allR$, // Set of all tested Idle-Busy ratios
$allV$, // Set of all tested traveling speeds
$allW$, // Set of all tested waiting time tolerances
$vehSamp = 20$, // Number of sampled sets of positions for drivers.
$pasSamp = 500$, // Number of sampled passengers per test.
Result: pl^s_{od} // Loss probabilities.
$auxPL \leftarrow$ 5-D array of zeros;
for each $idxF \in \{1,...,|allF|\}$ **do**
  $f \leftarrow allF(idxF)$;
  for each $idxR \in \{1,...,|allR|\}$ **do**
    $r \leftarrow allR(idxR)$;
    $f_{Idle} \leftarrow f * (1-r)$;
    $f_{Busy} \leftarrow f * r$;
    for each $idxP \in \{1,...,\text{vehSamp}\}$ **do**
      Initialize empty object VEH with properties $.orig, .dest, .curr, .busy$;
      Sample f_{Idle} Idle drivers into object VEH;
      Append a sample of f_{Busy} Busy drivers into object VEH;
      Sample a set of $pasSamp$ OD-pair into an object PAS;
      for each $idxV \in \{1,...,|allV|\}$ **do**
        $v_o \leftarrow allV(idxV)$;
        for each $idxW \in \{1,...,|allW|\}$ **do**
          $\omega \leftarrow allW(idxW)$;
          match $\leftarrow matdis(VEH, PAS, v_o, \omega)$; // Check match constraints
          /* Match is a vector with $pasSamp$ logical elements
            - 0 indicates that a passenger is NOT LOST.
            - 1 indicates that a passenger is LOST. */
          $auxPL(idxF, idxR, idxP, idxV, idxW) \leftarrow mean$(match);
        end
      end
    end
  end
end
$pl^s_{od} \leftarrow$ Average of $auxPL$ over the 3rd dimension ; // over $vehSamp$ tests.
return pl^s_{od} ;

Algorithm 1: Monte Carlo algorithm pseudo-code.

Data: $allDist$, // Matrix with the shortest distance for every OD-pair.
$detour$, // Detour tolerance.
function $matdis(VEH, PAS, v_o, \omega)$**:**
match ← vector with $|PAS|$ elements all equal to 'true';
for each $pas \in PAS$ **do**
$k \leftarrow$ true;
$veh \leftarrow 0$;
while k is true **&&** $veh \leq |VEH|$ **do**
$veh \leftarrow veh + 1$;
if $allDist(VEH(veh).curr, PAS(pas).origin)/v \leq \omega$ **then**
if $VEH(veh).busy == true$ **then**
match(pas) ← false;
$k \leftarrow$ false;
else
$dist01 \leftarrow allDist(VEH(veh).orig, VEH(veh).dest) * (1 + detour)$;
$dist02 \leftarrow allDist(pas.orig, pas.dest) * (1 + detour)$;
$dist11 \leftarrow allDist(VEH(veh).orig, VEH(veh).curr) +$
$allDist(VEH(veh).curr, pas.orig) +$
$allDist(pas.orig, VEH(veh).dest)$;
$dist12 \leftarrow allDist(pas.orig, VEH(veh).dest) +$
$allDist(VEH(veh).dest, pas.dest)$;
$dist21 \leftarrow allDist(VEH(veh).orig, VEH(veh).curr) +$
$allDist(VEH(veh).curr, pas.orig) + allDist(pas.orig, pas.dest) +$
$allDist(pas.dest, VEH(veh).dest)$;
$dist22 \leftarrow allDist(pas.orig, pas.dest)$;
// Test ridesplitting (two possible sequences).
if $dist11 \leq dist01$ **&&** $dist12 \leq dist02$ **then**
match(passenger) ← false;
$k \leftarrow$ false;
end
if $dist21 \leq dist01$ **&&** $dist22 \leq dist02$ **then**
match(passenger) ← false;
$k \leftarrow$ false;
end
end
end
end
if *vehicle* pick-up time ≤ ω && respecting possible detour constraints **then**
end
end
return match

Algorithm 2: Passenger-Driver matching test pseudo-code.

C Loss probability function estimation

As presented earlier, Equation [3.14] is a function of the available fleet size $n_{od}^{av}(t)$, the regional average speed $v_o(t)$, the waiting time tolerance ω, and the ratio of Idle-Busy drivers $\rho^s = n^I(t)/n_{od}^{av}(t)$. Hence, we can linearize Equation [3.14] into Equation [C.1], which allows us to estimate the coefficients to its equivalent linear regression model in Equation [C.2] (written using Wilkinson notation) using the least-square fit.

$$\log(-\log(pl_{od}^s)) = \log\gamma_0 + \gamma_1 \log n_{od}^{av} + \gamma_2 \log v_o + \gamma_3 \log\omega + \gamma_4 \log\rho^s \tag{C.1}$$

$$Y \sim I + n_{od}^{av} + v_o + \omega + \rho^s \tag{C.2}$$

For the sake of illustration, Figure C.1 shows an instance of the fitted function in comparison with the simulated data of the two-region setting.

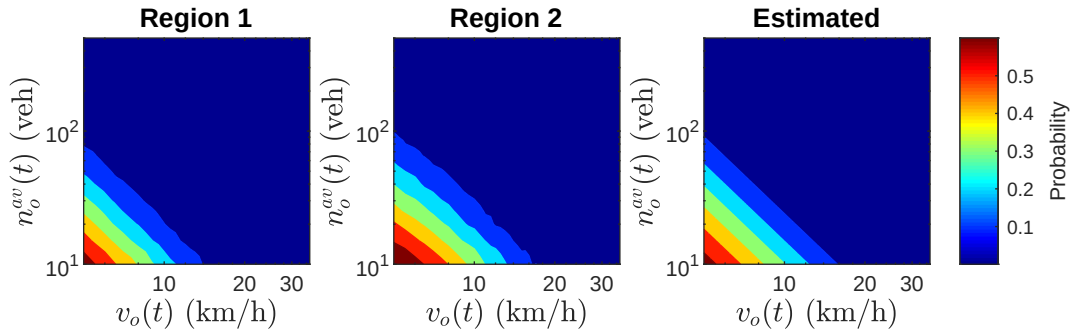


Figure C.1: Estimated loss probability for passengers (both axes in log-scale): (a) Monte Carlo Simulation for region 1, (b) Monte Carlo Simulation for region 2, (c) Fitted equation results. Where $n_o^{av}(t)$ is the fleet of available vehicles at time t, while $v_o(t)$ is the instantaneous speed in the region o at time t. The values for waiting time tolerance ω and the ratio of Idle-Busy drivers rho^s were fixed at 10 minutes and 1, respectively.

Bibliography

Aalipour, A., Kebriaei, H., & Ramezani, M. (2019). Analytical optimal solution of perimeter traffic flow control based on mfd dynamics: a pontryagin's maximum principle approach. *IEEE Transactions on Intelligent Transportation Systems*, *20*(9), 3224–3234. https://doi.org/10.1109/TITS.2018.2873104

Aboudolas, K., & Geroliminis, N. (2013). Perimeter and boundary flow control in multi-reservoir heterogeneous networks. *Transportation Research Part B: Methodological*, *55*, 265–281. https://doi.org/10.1016/j.trb.2013.07.003

Afeche, P., Liu, Z., & Maglaras, C. (2018). Ride-hailing networks with strategic drivers: the impact of platform control capabilities on performance. *Columbia Business School Research Paper No. 18–19*. https://doi.org/10.2139/ssrn.3120544

Agatz, N., Erera, A., Savelsbergh, M., & Wang, X. (2012). Optimization for dynamic ride-sharing: a review. *European Journal of Operational Research*, *223*, 295–303. https://doi.org/10.1016/j.ejor.2012.05.028

Alemi, F., Circella, G., Handy, S., & Mokhtarian, P. (2018). What influences travelers to use Uber? exploring the factors affecting the adoption of on-demand ride services in California. *Travel Behaviour and Society*, *13*, 88–104. https://doi.org/10.1016/j.tbs.2018.06.002

Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., & Rus, D. (2017). On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences USA*, *114*(3), 462–467. https://doi.org/10.1073/pnas.1611675114

Ampountolas, K., Zheng, N., & Geroliminis, N. (2017). Macroscopic modelling and robust control of bi-modal multi-region urban road networks. *Transportation Research Part B: Methodological*, *104*, 616–637. https://doi.org/10.1016/j.trb.2017.05.007

Arnott, R. (2013). A bathtub model of downtown traffic congestion. *Journal of Urban Economics*, *76*, 110–121. https://doi.org/10.1016/j.jue.2013.01.001

Band Jornalismo. (2023, March 9). *Prefeitura lança o aplicativo mobizapsp para transportes na capital*. Retrieved April 19, 2023, from https://www.band.uol.com.br/band-fm/noticias/prefeitura-lanca-o-aplicativo-mobizapsp-para-transportes-na-capital-16588043

Bassolas, A., Gallotti, R., Lamanna, F., Lenormand, M., & Ramasco, J. J. (2020). Scaling in the recovery of urban transportation systems from massive events. *Scientific Reports*, *10*, 2746. https://doi.org/10.1038/s41598-020-59576-1

Batista, S. F. A., & Leclercq, L. (2019). Regional dynamic traffic assignment framework for macroscopic fundamental diagram multi-regions models. *Transportation Science*, *53*(6), 1563–1590. https://doi.org/10.1287/trsc.2019.0921

Batista, S., Leclercq, L., & Geroliminis, N. (2019). Estimation of regional trip length distributions for the calibration of the aggregated network traffic models. *Transportation Research Part B: Methodological*, *122*, 192–217. https://doi.org/10.1016/j.trb.2019.02.009

Bellocchi, L., & Geroliminis, N. (2020). Unraveling reaction-diffusion-like dynamics in urban congestion propagation: insights from a large-scale road network. *Scientific Reports*, *10*, 4876. https://doi.org/10.1038/s41598-020-61486-1

Beojone, C. V., & Geroliminis, N. (2019). Accessing shared mobility impacts on urban traffic networks through discrete event simulations. *8th Symposium of the European Association for Research in Transportation (hEART 2019)*.

Beojone, C. V., & Geroliminis, N. (2020a). Macroscopic modeling of ride-sourcing services. *20th Swiss Transport Research Conference (STRC 2020)*.

Beojone, C. V., & Geroliminis, N. (2020b). Towards sustainable ride-sourcing services: a simulation study on the effects of congestion, fleet size and willingness to share. *99th Transportation Research Board Annual Meeting (TRB 2020)*.

Beojone, C. V., & Geroliminis, N. (2021a). Macroscopic ride-sourcing model for fleet rebalancing. *100th Transportation Research Board Annual Meeting (TRB 2021)*.

Beojone, C. V., & Geroliminis, N. (2021b). On the inefficiency of ride-sourcing services towards urban congestion. *Transportation Research Part C: Emerging Technologies*, *124*, 102890. https://doi.org/10.1016/j.trc.2020.102890

Beojone, C. V., & Geroliminis, N. (2021c). A path to take passengers from single to shared rides: a study on ridesplitting. *9th Symposium of the European Association for Research in Transportation (hEART 2020)*.

Beojone, C. V., & Geroliminis, N. (2021d). Repositioning idle vehicles in ridesplitting operations using pricing. *21st Swiss Transport Research Conference (STRC 2021)*.

Beojone, C. V., & Geroliminis, N. (2022a). A dynamic multi-region mfd model for ride-sourcing systems with ridesplitting. *101st Transportation Research Board Annual Meeting (TRB2022)*.

Beojone, C. V., & Geroliminis, N. (2022b). An optimized driver repositioning strategy in ridesplitting with earning estimates: a two-layer dynamic model and control. *10th Symposium of the European Association for Research in Transportation (hEART 2022)*.

Beojone, C. V., & Geroliminis, N. (2023a). A dynamic multi-region mfd model for ride-sourcing systems with ridesplitting. *Transportation Research Part B: Methodological, (Under Review)*.

Beojone, C. V., & Geroliminis, N. (2023b). Providing a revenue-forecasting information scheme as an incentive to relocate compliant ride-sourcing drivers. *102nd Transportation Research Board (TRB) Annual Meeting*, 1–15.

Beojone, C. V., & Geroliminis, N. (2023c). Relocation incentives for ride-sourcing drivers with path-oriented revenue forecasting based on a markov chain model. *Transportation Research Part C: Emerging Technologies, (Under Review)*.

Beojone, C. V., Geroliminis, N., & Yin, Y. (2022). Repositioning ridesplitting vehicles through pricing: a two-region simulated study. *11th Triennial Symposium on Transportation Analysis (XI TRISTAN)*.

Beojone, C. V., Zhu, P., Sırmatel, I. İ., & Geroliminis, N. (2023). A hierarchical control framework for vehicle repositioning in ride-hailing systems. *25th International Symposium on Transportation and Traffic Theory (accepted for full paper submission)*.

Berbeglia, G., Cordeau, J.-F., & Laporte, G. (2010). Dynamic pickup and delivery problems. *European Journal of Operational Research*, *202*, 8–15. https://doi.org/10.1016/j.ejor.2009.04.024

Bongiovanni, C., Kaspi, M., & Geroliminis, N. (2019). The electric autonomous dial-a-ride problem. *Transportation Research Part B*, *122*, 436–456. https://doi.org/10.1016/j.trb.2019.03.004

Castillo, J., Knoepfle, D., & Weyl, E. (2018). Surge pricing solves the wild goose chase. *Social Science Research Network*, *2890666*. https://doi.org/10.2139/ssrn.2890666

Chen, X. (, He, X., Xiong, C., Zhu, Z., & Zhang, L. (2019). A bayesian stochastic kriging optimization model dealing with heteroscedastic simulation noise for freeway traffic management. *Transportation Science*, *53*(2), 545–565. https://doi.org/10.1287/trsc.2018.0819

Çolak, S., Lima, A., & González, M. C. (2016). Understanding congested travel in urban areas. *Nature Communications*, *7*(10793). https://doi.org/10.1038/ncomms10793

Contreras, S. D., & Paz, A. (2018). The effects of ride-hailing companies on the taxicab industry in Las Vegas, Nevada. *Transportation Research Part A, 115*, 63–70. https://doi.org/10.1016/j.tra.2017.11.008

Cortés, C. E., Matamala, M., & Contardo, C. (2010). The pickup and delivery problem with transfers: formulation and branch-and-cut solution method. *European Journal of Operational Research*, *200*, 711–724. https://doi.org/10.1016/j.ejor.2009.01.022

Cox III, J., & Schleier, J. (2010). *Theory of constraints handbook*. McGraw-Hill.

Csikós, A., Charalambous, T., Farhadi, H., Kulcsár, B., & Wymeersch, H. (2017). Network traffic flow optimization under performance constraints. *Transportation Research Part C: Emerging Technologies*, *83*, 120–133. https://doi.org/10.1016/j.trc.2017.08.002

Daganzo, C. F. (2007). Urban gridlock: macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological*, *41*, 49–62. https://doi.org/10.1016/j.trb.2006.03.001

Daganzo, C. F. (2010). *Public Transportation Systems:Basic Principles of System Design,Operations Planning and Real-TimeControl*. UC Berkeley Course Notes Series.

Daganzo, C. F., & Lehe, L. J. (2015). Distance-dependent congestion pricing for downtown zones. *Transportation Research Part B: Methodological*, *75*, 89–99. https://doi.org/10.1016/j.trb.2015.02.010

Daganzo, C., & Ouyang, Y. (2019). *Public transportation systems: principles of system design, operations planning and real-time control*. World Scientific Publishing Co. https://doi.org/10.1142/10553

Ding, H., Zhang, Y., Zheng, X., Yuan, H., & Zhang, W. (2018). Hybrid perimeter control for two-region urban cities with different states. *IEEE Transactions on Control Systems Technology*, *26*(6), 2049–2062. https://doi.org/10.1109/TCST.2017.2746061

Dong, Y., Wang, S., Li, L., & Zhang, Z. (2018). An empirical study on travel patterns of internet based ride-sharing. *Transportation Research Part C: Emerging Technologies*, *86*, 1–22. https://doi.org/10.1016/j.trc.2017.10.022

Douglas, G. W. (1972). Price regulation and optimal service standards: the taxicab industry. *Journal of Transport Economics and Policy*, *6*(2), 116–127.

Durham, J. W., Carli, R., Frasca, P., & Bullo, F. (2012). Discrete partitioning and coverage control for gossiping robots. *IEEE Transactions on Robotics*, *28*(2), 364–378.

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). https://doi.org/10.1126/sciadv.aau2670

Erwig, M. (2000). The graph voronoi diagram with applications. *Networks*, *36*, 156–163.

Fosgerau, M. (2009). The marginal social cost of headway for a scheduled service. *Transportation Research Part B: Methodological*, *43*(8), 813–820. https://doi.org/10.1016/j.trb.2009.02.006

Fosgerau, M. (2015). Congestion in the bathtub. *Economics of Transportation*, *4*(4), 241–255. https://doi.org/10.1016/j.ecotra.2015.08.001

Fu, H., Liu, N., & Hu, G. (2017). Hierarchical perimeter control with guaranteed stability for dynamically coupled heterogeneous urban traffic. *Transportation Research Part C: Emerging Technologies*, *83*, 18–38. https://doi.org/10.1016/j.trc.2017.07.007

Fu, H., Wang, Y., Tang, X., Zheng, N., & Geroliminis, N. (2020). Empirical analysis of large-scale multimodal traffic with multi-sensor data. *Transportation Research Part C: Emerging Technologies*, *118*, 102725. https://doi.org/10.1016/j.trc.2020.102725

Furuhata, M., Dessouky, M., Ordóñez, F., Brunet, M.-E., Wang, X., & Koenig, S. (2013). Ridesharing: the state-of-the-art and future directions. *Transportation Research Part B*, *57*, 28–46. https://doi.org/10.1016/j.trb.2013.08.012

Gama, P. (2022). Insatisfeitos com uber e 99, motoristas criam aplicativos alternativos. *UOL Notícias*. Retrieved April 19, 2023, from https://www.uol.com.br/carros/colunas/paula-gama/2022/12/07/insatisfeitos-com-uber-e-99-motoristas-de-criam-aplicativos-alternativos.htm

Geroliminis, N. (2015). Cruising-for-parking in congested cities with an mfd representation. *Economics of Transportation*, *4*(3), 156–165. https://doi.org/10.1016/j.ecotra.2015.04.001

Geroliminis, N., & Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. *Transportation Research Part B*, *42*, 759–770. https://doi.org/10.1016/j.trb.2008.02.002

Geroliminis, N., & Sun, J. (2011). Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B: Methodological*, *45*(3), 605–617. https://doi.org/10.1016/j.trb.2010.11.004

Geroliminis, N., Zheng, N., & Ampountolas, K. (2014). A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. *Transportation Research Part C: Emerging Technologies*, *42*, 168–181. https://doi.org/10.1016/j.trc.2014.03.004

Haddad, J. (2015). Robust constrained control of uncertain macroscopic fundamental diagram networks [Special Issue on International Symposium on Transportation and Traffic Theory]. *Transportation Research Part C: Emerging Technologies*, *59*, 323–339. https://doi.org/10.1016/j.trc.2015.05.014

Haddad, J. (2017). Optimal perimeter control synthesis for two urban regions with aggregate boundary queue dynamics. *Transportation Research Part B: Methodological*, *96*, 1–25. https://doi.org/10.1016/j.trb.2016.10.016

Haddad, J., & Mirkin, B. (2016). Adaptive perimeter traffic control of urban road networks based on mfd model with time delays. *International Journal of Robust and Nonlinear Control*, *26*(6), 1267–1285. https://doi.org/10.1002/rnc.3502

Haddad, J., Mirkin, B., & Assor, K. (2021). Traffic flow modeling and feedback control for future low-altitude air city transport: an mfd-based approach. *Transportation Research Part C: Emerging Technologies*, *133*, 103380. https://doi.org/10.1016/j.trc.2021.103380

Haddad, J., & Shraiber, A. (2014). Robust perimeter control design for an urban region. *Transportation Research Part B: Methodological*, *68*, 315–332. https://doi.org/10.1016/j.trb.2014.06.010

Haddad, J., & Zheng, Z. (2020). Adaptive perimeter control for multi-region accumulation-based models with state delays [Advances in Network Macroscopic Fundamental Diagram (NMFD) Research]. *Transportation Research Part B: Methodological*, *137*, 133–153. https://doi.org/10.1016/j.trb.2018.05.019

Hall, J. D., Palsson, C., & Price, J. (2018). Is uber a substitute or complement for public transit? *Journal of Urban Economics*, *108*, 36–50. https://doi.org/10.1016/j.jue.2018.09.003

Hamedmoghadam, H., Ramezani, M., & Saberi, M. (2019). Revealing latent characteristics of mobility networks with coarse-graining. *Scientific Reports*, *9*, 7545. https://doi.org/10.1038/s41598-019-44005-9

Hanna, J. P., Albert, M., Chen, D., & Stone, P. (2016). Minimum cost matching for autonomous carsharing [9th IFAC Symposium on Intelligent Autonomous Vehicles IAV 2016]. *IFAC-PapersOnLine*, *49*(15), 254–259. https://doi.org/10.1016/j.ifacol.2016.07.757

Hans, E., Chiabaut, N., & Leclercq, L. (2014). Clustering approach for assessing the travel time variability of arterials. *Transportation Research Record*, *2422*(1), 42–49. https://doi.org/10.3141/2422-05

Hanson, S., & Giuliano, G. (2017). *The geography of urban transportation* (4th). Guilford Press.

He, F., & Shen, Z.-J. M. (2015). Modeling taxi services with smartphone-based e-hailing applications. *Transportation Research Part C*, *58*, 93–106. https://doi.org/10.1016/j.trc.2015.06.023

Hensher, D. A., & Rose, J. M. (2007). Development of commuter and non-commuter mode choice models for assessment of new public transport infrastructure projects: a case study. *Transportation Research Part A*, *41*, 428–443. https://doi.org/10.1016/j.tra.2006.09.006

Ho, S. C., Szeto, W. Y., Kuo, Y.-H., Leung, J. M. Y., Petering, M., & Tou, W. H., Terence. (2018). A survey of dial-a-ride problems: literature review and recent developments. *Transportation Research Part B, 111*, 395–421. https://doi.org/10.1016/j.trb.2018.02.001

Hosni, H., Naoum-Sawaya, J., & Artail, H. (2014). The shared-taxi problem: formulation and solution methods. *Transportation Research Part B, 70*, 303–318. https://doi.org/10.1016/j.trb.2014.09.011

Ingole, D., Mariotte, G., & Leclercq, L. (2020). Perimeter gating control and citywide dynamic user equilibrium: a macroscopic modeling framework. *Transportation Research Part C: Emerging Technologies, 111*, 22–49. https://doi.org/10.1016/j.trc.2019.11.016

Ingolfsson, A. (2005). Modeling the M(t)/M/s(t) queue with an exhaustive discipline. *Working paper*. http://www.bus.ualberta.ca/aingolfsson/working%5C_papers.htm

Ingolfsson, A., Akhmetshina, E., Budge, S., Li, Y., & Wu, X. (2007). A survey and experimental comparison of service-level-approximation methods for nonstationary M (t)/M/s(t) queueing systems with exhaustive discipline. *INFORMS journal on computing, 19*(2), 201–214. https://doi.org/10.1287/ijoc.1050.0157

Ji, Y., Luo, J., & Geroliminis, N. (2014). Empirical observations of congestion propagation and dynamic partitioning with probe data for large-scale systems. *Transportation Research Record: Journal of the Transportation Research Board, 2422*, 1–11. https://doi.org/10.3141/2422-01

Jin, S. T., Kong, H., Wu, R., & Sui, D. Z. (2018). Ridesourcing, the sharing economy, and the future of cities. *Cities, 76*, 96–104. https://doi.org/10.1016/j.cities.2018.01.012

Jin, W.-L. (2020). Generalized bathtub model of network trip flows. *Transportation Research Part B: Methodological, 136*, 138–157. https://doi.org/10.1016/j.trb.2020.04.002

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. https://doi.org/10.1111/mice.12157

Ke, J., Li, X., Yang, H., & Yin, Y. (2021). Pareto-efficient solutions and regulations of congested ride-sourcing markets with heterogeneous demand and supply. *Social Science Research Network, 3773481*. https://doi.org/10.2139/ssrn.3773481

Ker, J. (2023). 'Uber' lançado pela prefeitura de sp sofre com falta de carros e greve em seu primeiro dia. *O Estado de São Paulo Conteúdo*. Retrieved April 19, 2023, from https://www.estadao.com.br/sao-paulo/uber-lancado-pela-prefeitura-de-sp-sofre-com-falta-de-carros-e-greve-em-seu-primeiro-dia/

Kouvelas, A., Saeedmanesh, M., & Geroliminis, N. (2017). Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. *Transportation Research Part B: Methodological, 96*, 26–45. https://doi.org/10.1016/j.trb.2016.10.011

Kuhn, H. W. (2005). The hungarian method for the assignment problem. *Naval Research Logistics (NRL), 52*(1), 7–21. https://doi.org/10.1002/nav.20053

Kumarage, S., Yildirimoglu, M., & Zheng, Z. (2023). Demand and state estimation for perimeter control in large-scale urban networks. *Transportation Research Part C: Emerging Technologies, 153*, 104184. https://doi.org/10.1016/j.trc.2023.104184

Lamotte, R., de Palma, A., & Geroliminis, N. (2017). On the use of reservation-based autonomous vehicles for demand management. *Transportation Research Part B: Methodological*, *99*, 205–227. https://doi.org/10.1016/j.trb.2017.01.003

Lamotte, R., & Geroliminis, N. (2018). The morning commute in urban areas with heterogeneous trip lengths [TRB:ISTTT-22]. *Transportation Research Part B: Methodological*, *117*, 794–810. https://doi.org/10.1016/j.trb.2017.08.023

Lamotte, R., Murashkin, M., Kouvelas, A., & Geroliminis, N. (2018). Dynamic modeling of trip completion rate in urban areas with mfd representations. *TRB Annual Meeting*, 18–06192. https://doi.org/10.3929/ethz-b-000275987

Larson, R., & Odoni, A. (1981). *Urban operations research*. Prentice-Hall. https://web.mit.edu/urban%7B%5C_%7Dor%7B%5C_%7Dbook/www/book/

Lavieri, P. S., & Bhat, C. R. (2019). Investigating objective and subjective factors influencing the adoption, frequency, and characteristics of ride-hailing trips. *Transportation Research Part C: Emerging Technologies*, *105*, 100–125. https://doi.org/10.1016/j.trc.2019.05.037

Lawphongpanich, S., & Yin, Y. (2012). Nonlinear pricing on transportation networks [Special issue on Optimization in Public Transport+ISTT2011]. *Transportation Research Part C: Emerging Technologies*, *20*(1), 218–235. https://doi.org/10.1016/j.trc.2011.05.010

Leclercq, L., & Paipuri, M. (2019). Macroscopic traffic dynamics under fast-varying demand. *Transportation Science*, *53*(6), 1526–1545. https://doi.org/10.1287/trsc.2019.0908

Lee, D.-H., Wang, H., Cheu, R. L., & Teo, S. H. (2004). A taxi dispatch system based on current demands and real-time traffic conditions. *Transportation Research Record: Journal of the Transportation Research Board*, *1882*(1), 193–200. https://doi.org/10.3141/1882-23

Li, W., Pu, Z., Li, Y., & Ban, X. ((2019). Characterization of ridesplitting based on observed data: a case study of Chengdu, China. *Transportation Research Part C*, *100*, 330–353. https://doi.org/10.1016/j.trc.2019.01.030

Li, Y., Yildirimoglu, M., & Ramezani, M. (2021). Robust perimeter control with cordon queues and heterogeneous transfer flows. *Transportation Research Part C: Emerging Technologies*, *126*, 103043. https://doi.org/10.1016/j.trc.2021.103043

Liu, Y., & Samaranayake, S. (2020). Proactive rebalancing and speed-up techniques for on-demand high capacity ridesourcing services. *IEEE Transactions on Intelligent Transportation Systems*, 1–8. https://doi.org/10.1109/TITS.2020.3016128

Loder, A., Ambühl, L., Menendez, M., & Axhausen, K. W. (2019). Understanding traffic capacity of urban networks. *Scientific Reports*, *9*, 16283. https://doi.org/10.1038/s41598-019-51539-5

Long, J., Tan, W., Szeto, W. Y., & Li, Y. (2018). Ride-sharing with travel time uncertainty. *Transportation Research Part B*, *118*, 143–171. https://doi.org/10.1016/j.trb.2018.10.004

Lopez, C., Leclercq, L., Krishnakumari, P., Chiabaut, N., & van Lint, H. (2017). Revealing the day-to-day regularity of urban congestion patterns with 3d speed maps. *Scientific Reports*, *7*, 14029. https://doi.org/10.1038/s41598-017-14237-8

Lu, A., Frazier, P., & Kislev, O. (2018). Surge pricing moves uber's driver partners. *Social Science Research Network*, *3180246*. https://doi.org/10.2139/ssrn.3180246

Mariotte, G., Leclercq, L., & Laval, J. A. (2017). Macroscopic urban dynamics: analytical and numerical comparisons of existing models. *Transportation Research Part B: Methodological*, *101*, 245–267. https://doi.org/https://doi.org/10.1016/j.trb.2017.04.002

Martinez, L. M., Correia, G. H. A., & Viegas, J. M. (2015). An agent-based simulation model to assess the impacts of introducing a shared-taxi system: an application to Lisbon (Portugal). *Journal of Advanced Transportation*, *49*(3), 475–495. https://doi.org/10.1002/atr.1283

Masmoudi, M. A., Hosny, M., Demir, E., Genikomsakis, K. N., & Cheikhrouhou, N. (2018). The dial-a-ride problem with electric vehicles and battery swapping. *Transportation Research Part E*, *118*, 392–420. https://doi.org/10.1016/j.tre.2018.08.005

Menelaou, C., Kolios, P., Timotheou, S., Panayiotou, C., & Polycarpou, M. (2017). Controlling road congestion via a low-complexity route reservation approach. *Transportation Research Part C: Emerging Technologies*, *81*, 118–136. https://doi.org/10.1016/j.trc.2017.05.005

Menelaou, C., Timotheou, S., Kolios, P., Panayiotou, C. G., & Polycarpou, M. M. (2019). Minimizing traffic congestion through continuous-time route reservations with travel time predictions. *IEEE Transactions on Intelligent Vehicles*, *4*(1), 141–153. https://doi.org/10.1109/TIV.2018.2886684

Mohajerpoor, R., Saberi, M., Vu, H. L., Garoni, T. M., & Ramezani, M. (2020). H_∞ robust perimeter flow control in urban networks with partial information feedback [Advances in Network Macroscopic Fundamental Diagram (NMFD) Research]. *Transportation Research Part B: Methodological*, *137*, 47–73. https://doi.org/10.1016/j.trb.2019.03.010

Molenbruch, Y., Braekers, K., & Caris, A. (2017). Typology and literature review for dial-a-ride problems. *Annals of Operations Research*, *259*, 295–325. https://doi.org/10.1007/s10479-017-2525-0

Munkres, J. (1957). Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, *5*(1), 32–38.

Murashkin, M. (2021). The influence of trip length distribution on urban traffic in network-level models. *EPFL PhD thesis*, 124 pages. https://doi.org/10.5075/epfl-thesis-9638

Narayanan, S., Chaniotakis, E., & Antoniou, C. (2020). Shared autonomous vehicle services: a comprehensive review. *Transportation Research Part C: Emerging Technologies*, *111*, 255–293. https://doi.org/10.1016/j.trc.2019.12.008

Ni, W., & Cassidy, M. (2020). City-wide traffic control: modeling impacts of cordon queues. *Transportation Research Part C: Emerging Technologies*, *113*, 164–175. https://doi.org/10.1016/j.trc.2019.04.024

Nourinejad, M., & Ramezani, M. (2020). Ride-sourcing modeling and pricing in non-equilibrium two-sided markets [23rd International Symposium on Transportation and Traffic Theory (ISTTT 23)]. *Transportation Research Part B: Methodological*, *132*, 340–357. https://doi.org/10.1016/j.trb.2019.05.019

Nourinejad, M., & Roorda, M. J. (2016). Agent based model for dynamic ridesharing. *Transportation Research Part C*, *64*, 117–132. https://doi.org/10.1016/j.trc.2015.07.016

Ohnemus, M., & Perl, A. (2016). Shared autonomous vehicles: catalyst of new mobility for the last mile? *Built Environment (1978-), 42*(4), 589–602. https://doi.org/10.2148/benv.42.4.589

Olmos, L. E., Çolak, S., Shafiei, S., Saberi, M., & González, M. C. (2018). Macroscopic dynamics and the collapse of urban traffic. *Proceedings of the National Academy of Sciences USA, 115*(50), 12654–12661. https://doi.org/10.1073/pnas.1800474115

Owen, S. H., & Daskin, M. S. (1998). Strategic facility location: a review. *European Journal of Operational Research, 111*, 423–447. https://doi.org/10.1016/S0377-2217(98)00186-6

Páez, A., Scott, D. M., & Morency, C. (2012). Measuring accessibility: positive and normative implementations of various accessibility indicators. *Journal of Transport Geography, 25*, 141–153. https://doi.org/10.1016/j.jtrangeo.2012.03.016

Paipuri, M., & Leclercq, L. (2020). Bi-modal macroscopic traffic dynamics in a single region. *Transportation Research Part B: Methodological, 133*, 257–290. https://doi.org/https://doi.org/10.1016/j.trb.2020.01.007

Pentico, D. W. (2007). Assignment problems: a golden anniversary survey. *European Journal of Operational Research, 176*(2), 774–793. https://doi.org/10.1016/j.ejor.2005.09.014

Powell, J. W., Huang, Y., Bastani, F., & Ji, M. (2011). Towards reducing taxicab cruising time using spatio-temporal profitability maps. In D. Pfoser, Y. Tao, K. Mouratidis, M. A. Nascimento, M. Mokbel, S. Shekhar, & Y. Huang (Eds.), *Advances in spatial and temporal databases* (pp. 242–260). Springer Berlin Heidelberg.

Ramezani, M., Haddad, J., & Geroliminis, N. (2015). Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. *Transportation Research Part B: Methodological, 74*, 1–19. https://doi.org/10.1016/j.trb.2014.12.010

Ramezani, M., & Nourinejad, M. (2018). Dynamic modeling and control of taxi services in large-scale urban networks: a macroscopic approach. *Transportation Research Part C, 94*, 203–219. https://doi.org/10.1016/j.trc.2017.08.011

Ramezani, M., & Valadkhani, A. H. (2023). Dynamic ride-sourcing systems for city-scale networks - part i: matching design and model formulation and validation. *Transportation Research Part C: Emerging Technologies, 152*, 104158. https://doi.org/10.1016/j.trc.2023.104158

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. https://doi.org/10.1016/j.tranpol.2015.10.004

Riascos, A. P., & Mateos, J. L. (2020). Networks and long-range mobility in cities: a study of more than one billion taxi trips in new york city. *Scientific Reports, 10*, 4022. https://doi.org/10.1038/s41598-020-60875-w

Rogers, B. (2017). The social costs of Uber. *The University of Chicago Law Review Online, 82*(1), Art. 6.

Roughgarden, T. (2005). *Selfish routing and the price of anarchy*. MIT Press.

Sadeghi, A., & Smith, S. L. (2019). On re-balancing self-interested agents in ride-sourcing transportation networks. *2019 IEEE 58th Conference on Decision and Control (CDC)*, 5119–5125. https://doi.org/10.1109/CDC40024.2019.9030043

Saedi, R., Saeedmanesh, M., Zockaie, A., Saberi, M., Geroliminis, N., & Mahmassani, H. S. (2020). Estimating network travel time reliability with network partitioning. *Transportation Research Part C: Emerging Technologies*, *112*, 46–61. https://doi.org/10.1016/j.trc.2020.01.013

Saeedmanesh, M., & Geroliminis, N. (2016). Clustering of heterogeneous networks with directional flows based on "snake" similarities. *Transportation Research Part B: Methodological*, *91*, 250–269. https://doi.org/10.1016/j.trb.2016.05.008

Safadi, Y., Fu, R., Quan, Q., & Haddad, J. (2023). Macroscopic fundamental diagrams for low-altitude air city transport. *Transportation Research Part C: Emerging Technologies*, *152*, 104141. https://doi.org/10.1016/j.trc.2023.104141

Saffari, E., Yildirimoglu, M., & Hickman, M. (2022). Data fusion for estimating macroscopic fundamental diagram in large-scale urban networks. *Transportation Research Part C: Emerging Technologies*, *137*, 103555. https://doi.org/10.1016/j.trc.2022.103555

Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S. H., & Ratti, C. (2014). Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy of Sciences USA*, *111*(37), 13290–13294. https://doi.org/10.1073/pnas.1403657111

Saw, V.-L., Chung, N. N., Quek, W. L., Pang, Y. E. I., & Chew, L. Y. (2019). Bus bunching as a synchronisation phenomenon. *Scientific Reports*, *9*, 6887. https://doi.org/10.1038/s41598-019-43310-7

Schwieterman, J., & Smith, C. S. (2018). Sharing the ride: a paired-trip analysis of UberPool and Chicago Transit Authority services in Chicago, Illinois. *Research in Transportation Economics*, *71*, 9–16. https://doi.org/10.1016/j.retrec.2018.10.003

Shou, Z., Di, X., Ye, J., Zhu, H., Zhang, H., & Hampshire, R. (2020). Optimal passenger-seeking policies on e-hailing platforms using markov decision process and imitation learning. *Transportation Research Part C: Emerging Technologies*, *111*, 91–113. https://doi.org/10.1016/j.trc.2019.12.005

Simonetto, A., Monteil, J., & Gambella, C. (2019). Real-time city-scale ridesharing via linear assignment problems. *Transportation Research Part C: Emerging Technologies*, *101*, 208–232. https://doi.org/10.1016/j.trc.2019.01.019

Sirmatel, I. I., & Geroliminis, N. (2018a). Economic model predictive control of large-scale urban road networks via perimeter control and regional route guidance. *IEEE Transactions on Intelligent Transportation Systems*, *19*(4), 1112–1121. https://doi.org/10.1109/TITS.2017.2716541

Sirmatel, I. I., & Geroliminis, N. (2018b). Mixed logical dynamical modeling and hybrid model predictive control of public transport operations. *Transportation Research Part B: Methodological*, *114*, 325–345. https://doi.org/10.1016/j.trb.2018.06.009

Sirmatel, I. I., & Geroliminis, N. (2020). Nonlinear moving horizon estimation for large-scale urban road networks. *IEEE Transactions on Intelligent Transportation Systems*, *21*(12), 4983–4994. https://doi.org/10.1109/TITS.2019.2946324

Sirmatel, I. I., & Geroliminis, N. (2021). Stabilization of city-scale road traffic networks via macroscopic fundamental diagram-based model predictive perimeter control. *Control Engineering Practice, 109*, 104750. https://doi.org/10.1016/j.conengprac.2021.104750

Sirmatel, I. I., Tsitsokas, D., Kouvelas, A., & Geroliminis, N. (2021). Modeling, estimation, and control in large-scale urban road networks with remaining travel distance dynamics [24th International Symposium on Transportation and Traffic Theory (ISTTT 24)]. *Transportation Research Part C: Emerging Technologies, In press*, 1–11.

Stiglic, M., Agatz, N., Savelsbergh, M., & Gradisar, M. (2016). Making dynamic ride-sharing work: the impact of driver and rider flexibility. *Transportation Research Part E, 91*, 190–207. https://doi.org/10.1016/j.tre.2016.04.010

Su, Q., & Wang, D. Z. (2019). Morning commute problem with supply management considering parking and ride-sourcing. *Transportation Research Part C: Emerging Technologies, 105*, 626–647. https://doi.org/10.1016/j.trc.2018.12.015

Tachet, R., Sagarra, O., Santi, P., Resta, G., Szell, M., Strogatz, S. H., & Ratti, C. (2017). Scaling law of urban ride sharing. *Scientific Reports, 7*, 42868. https://doi.org/10.1038/srep42868

Tirachini, A., & del Río, M. (2019). Ride-hailing in santiago de chile: users' characterisation and effects on travel behaviour. *Transport Policy, 82*, 46–57. https://doi.org/10.1016/j.tranpol.2019.07.008

Tirachini, A., & Gomez-Lobo, A. (2019). Does ride-hailing increase or decrease vehicle kilometers traveled (vkt)? a simulation approach for santiago de chile. *International Journal of Sustainable Transportation, 14*(3), 187–204. https://doi.org/10.1080/15568318.2018.1539146

Tisato, P. (1992). User cost minimisation and transport subsidy. *Economics Letters, 39*(2), 241–247. https://doi.org/10.1016/0165-1765(92)90297-C

Uber. (2019). Sample price from Union Square to SFO [Accessed: 2019-06-17].

Valadkhani, A. H., & Ramezani, M. (2023). Dynamic ride-sourcing systems for city-scale networks, part ii: proactive vehicle repositioning. *Transportation Research Part C: Emerging Technologies, 152*, 104159. https://doi.org/10.1016/j.trc.2023.104159

Vazifeh, M. M., Santi, P., Resta, G., Strogatz, S. H., & Ratti, C. (2018). Addressing the minimum fleet problem in on-demand urban mobility. *Nature, 557*, 534–538. https://doi.org/doi.org/10.1038/s41586-018-0095-1

Vickrey, W. (2020). Congestion in midtown manhattan in relation to marginal cost pricing. *Economics of Transportation, 21*, 100152. https://doi.org/10.1016/j.ecotra.2019.100152

Vickrey, W. S. (1969). Congestion theory and transport investment. *The American Economic Review, 59*(2), 251–260. https://doi.org/10.1016/0166-0462(75)90023-X

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. https://doi.org/10.1016/j.trc.2021.103088

Vinayak, P., Dias, F. F., Astroza, S., Bhat, C. R., Pendyala, R. M., & Garikapati, V. M. (2018). Accounting for multi-dimensional dependencies among decision-makers within a generalized model framework: an application to understanding shared mobility service usage levels. *Transport Policy, 72*, 129–137. https://doi.org/10.1016/j.tranpol.2018.09.013

Wallar, A., Van Der Zee, M., Alonso-Mora, J., & Rus, D. (2018). Vehicle rebalancing for mobility-on-demand systems with ride-sharing. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 4539–4546. https://doi.org/10.1109/IROS.2018.8593743

Wang, H., & Yang, H. (2019). Ridesourcing systems: a framework and review. *Transportation Research Part B: Methodological*, *129*, 122–155. https://doi.org/10.1016/j.trb.2019.07.009

Wei, B., Saberi, M., Zhang, F., Liu, W., & Waller, S. T. (2020). Modeling and managing ridesharing in a multi-modal network with an aggregate traffic representation: a doubly dynamical approach. *Transportation Research Part C: Emerging Technologies*, *117*, 102670. https://doi.org/10.1016/j.trc.2020.102670

Weihrich, H. (1982). The TOWS matrix – A tool for situational analysis. *Long Range Planning*, *15*(2), 54–66. https://doi.org/10.1016/0024-6301(82)90120-0

Wenzel, T., Rames, C., Kontou, E., & Henao, A. (2019). Travel and energy implications of ridesourcing service in austin, texas. *Transportation Research Part D: Transport and Environment*, *70*, 18–34. https://doi.org/10.1016/j.trd.2019.03.005

Wong, K. I., & Bell, M. G. H. (2006). The optimal dispatching of taxis under congestion: a rolling horizon approach. *Journal of Advanced Transportation*, *40*(2), 203–220. https://doi.org/10.1002/atr.5670400207

Xu, Z., Yin, Y., Chao, X., Zhu, H., & Ye, J. (2020). A generalized fluid model of ride-hailing systems. *Available at SSRN 3743112*.

Xu, Z., Yin, Y., & Ye, J. (2020). On the supply curve of ride-hailing systems [23rd International Symposium on Transportation and Traffic Theory (ISTTT 23)]. *Transportation Research Part B: Methodological*, *132*, 29–43. https://doi.org/10.1016/j.trb.2019.02.011

Xu, Z., Yin, Y., & Zha, L. (2017). Optimal parking provision for ride-sourcing services. *Transportation Research Part B: Methodological*, *105*, 559–578. https://doi.org/10.1016/j.trb.2017.10.003

Yang, H., Qin, X., Ke, J., & Ye, J. (2020). Optimizing matching time interval and matching radius in on-demand ride-sourcing markets. *Transportation Research Part B: Methodological*, *131*, 84–105. https://doi.org/10.1016/j.trb.2019.11.005

Yang, K., Zheng, N., & Menendez, M. (2018). Multi-scale perimeter control approach in a connected-vehicle environment [ISTTT22]. *Transportation Research Part C: Emerging Technologies*, *94*, 32–49. https://doi.org/10.1016/j.trc.2017.08.014

Yildirimoglu, M., & Geroliminis, N. (2014). Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams. *Transportation Research Part B: Methodological*, *70*, 186–200. https://doi.org/10.1016/j.trb.2014.09.002

Yildirimoglu, M., Sirmatel, I. I., & Geroliminis, N. (2018). Hierarchical control of heterogeneous large-scale urban road networks via path assignment and regional route guidance. *Transportation Research Part B: Methodological*, *118*, 106–123. https://doi.org/10.1016/j.trb.2018.10.007

Yin, R. K. (2018). *Case study research and applications*. Sage.

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. https://doi.org/10.1016/j.trb.2018.12.013

Zeng, W., Han, Y., Sun, W., & Xie, S. (2020). Exploring the ridesharing efficiency of taxi services. *IEEE Access*, *8*, 160396–160406. https://doi.org/10.1109/ACCESS.2020.3020861

Zha, L., Yin, Y., & Du, Y. (2018). Surge pricing and labor supply in the ride-sourcing market [TRB:ISTTT-22]. *Transportation Research Part B: Methodological*, *117*, 708–722. https://doi.org/10.1016/j.trb.2017.09.010

Zha, L., Yin, Y., & Yang, H. (2016). Economic analysis of ride-sourcing markets. *Transportation Research Part C*, *71*, 249–266. https://doi.org/10.1016/j.trc.2016.07.010

Zhang, K., Chen, H., Yao, S., Xu, L., Ge, J., Liu, X., & Nie, M. (2019). An efficiency paradox of uberization. *Social Science Research Network*, *3462912*. https://doi.org/10.2139/ssrn.3462912

Zhang, K., & Nie, M. (2019). To pool or not to pool: equilibrium, pricing and regulation. *Social Science Research Network*, *3497808*. https://doi.org/10.2139/ssrn.3497808

Zhang, K., & Nie, Y. ((2021). Inter-platform competition in a regulated ride-hail market with pooling. *Transportation Research Part E: Logistics and Transportation Review*, *151*, 102327. https://doi.org/10.1016/j.tre.2021.102327

Zhang, L., Zeng, G., Li, D., Huang, H.-J., Stanley, H. E., & Havlin, S. (2019). Scale-free resilience of real traffic jams. *Proceedings of the National Academy of Sciences USA*, *116*(18), 8673–8678. https://doi.org/10.1073/pnas.1814982116

Zhong, R., Chen, C., Huang, Y., Sumalee, A., Lam, W., & Xu, D. (2018). Robust perimeter control for two urban regions with macroscopic fundamental diagrams: a control-lyapunov function approach [TRB:ISTTT-22]. *Transportation Research Part B: Methodological*, *117*, 687–707. https://doi.org/10.1016/j.trb.2017.09.008

Zhou, X., Rong, H., Yang, C., Zhang, Q., Khezerlou, A. V., Zheng, H., Shafiq, Z., & Liu, A. X. (2020). Optimizing taxi driver profit efficiency: a spatial network-based markov decision process approach. *IEEE Transactions on Big Data*, *6*(1), 145–158. https://doi.org/10.1109/TBDATA.2018.2875524

Zhou, Z., De Schutter, B., Lin, S., & Xi, Y. (2017). Two-level hierarchical model-based predictive control for large-scale urban traffic networks. *IEEE Transactions on Control Systems Technology*, *25*(2), 496–508. https://doi.org/10.1109/TCST.2016.2572169

Zhu, P., Sirmatel, I. I., Ferrari-Trecate, G., & Geroliminis, N. (2022). Idle-vehicle rebalancing coverage control for ride-sourcing systems. *the European Control Conference 2022 (ECC22)*.

Caio Vitor Beojone

Doctoral assistant

GC C2 390 Station 18
1015 Lausanne
Switzerland
+41 78 804 19 09
+41 21 693 53 97
caio.beojone@epfl.ch
Google Scholar profile
⋆ 18.06.1992, Bauru, SP, Brazil

Research Interests

Transportation, shared mobility, on-demand transportation, traffic flow modeling, intelligent transportation systems.

Education

10.2018 – 07.2023 **Ph.D., civil engineering**, *Urban Transport Systems Laboratory (LUTS)*, *École Polytechnique Fédérale de Lausanne (EPFL)*
Thesis: Modeling and Optimization of Ridesplitting Operations. Supervisor: Prof. Nikolas Geroliminis

08.2015– 11.2017 **M.Sc., production engineering**, *Department of Production Engineering*, *São Paulo State University (UNESP)*
Dissertation: Performance and alternative scenarios evaluation on a SAMU using the stationary and nonstationary hypercube model. (text in Portuguese) Advisor: Prof. Regiane Máximo Siqueira

08.2010 – 10.2015 **B.Sc., production engineering**, *Department of Production Engineering*, *São Paulo State University (UNESP)*
Graduation work: Analysis of a SAMU using the hypercube queueing model with partial backup. (text in Portuguese) Advisor: Prof. Regiane Máximo Siqueira

Research Experience

03.2022 – 06.2022 **Visiting scholar**, *University of Michigan*
Topic: Could occupation-based non-linear pricing make ride-pooling a viable alternative during peak hours and improve social-welfare?
Advisor: Prof. Yafeng Yin.

10.2018 – 07.2023 **Doctoral assistant**, *École Polytechnique Fédérale de Lausanne*

2018–2022 **Master thesis supervisor**, *École Polytechnique Fédérale de Lausanne*
Arseni Fedosseev, Design of ride-hailing services for large-scale systems, 2020.
Lynn Fayed, Simulated Annealing Algorithm for Solving the Matching and Dispatching Problem in Dynamic Large-Scale Settings, 2020
Minru Wang. The role of ride-split revenue optimization on service level and traffic operation, 2021.

2019-2022 **Semester project supervisor**, *École Polytechnique Fédérale de Lausanne*
Andrea Alfred Vanoncini, Frederik Endras, *Planning ride-sharing systems under different passengers behaviours*, 2019.
Nicolas Salvadé, Louise Aoustin, *Efficiency of ridesplitting for different street network topologies*, 2019.
Simon Dayer, Samuel Masmejan, *Ride-sourcing in real life: Descriptive Analysis on Chicago's Transportation Network Companies data*, 2020.
Yannis Voet, *Data analysis Uber network Chicago*, 2020
Alfio Mosset, Nicolas Salvaé, Pablo Cattin, *The dynamic matching problem in large-scale ride-sourcing systems*, 2020.

08.2015–11.2017 **Master student**, *São Paulo State University (UNESP)*
Topic: Performance and alternative scenarios evaluation on a SAMU using the stationary and nonstationary hypercube model.

Teaching Experience

2019–2021 **Teaching assistant**, *École Polytechnique Fédérale de Lausanne*
Course: Transportation systems engineering I

2021 and 2022 **Guest lecture**
Course: Fundamentals of traffic operations and control, Lecture: Shared mobility and Mobility as a Service.

2017 **Teaching Assistant**, *São Paulo State University (UNESP)*
Course: Pesquisa Operacional II (Operations Research II)

Publications

Peer-Reviewed Journal Articles

[J3] C. V. Beojone and N. Geroliminis, "On the inefficiency of ride-sourcing services towards urban congestion," *Transportation research part C: emerging technologies*, vol. 124, p. 102 890, 2021

[J2] C. V. Beojone, R. Máximo de Souza, and A. P. Iannoni, "An efficient exact hypercube model with fully dedicated servers," *Transportation Science*, vol. 55, no. 1, pp. 222–237, 2021

[J1] C. V. Beojone and R. M. d. Souza, "Improving the shift-scheduling problem using non-stationary queueing models with local heuristic and genetic algorithm," *Pesquisa Operacional*, vol. 40, 2020

Journal Articles Under Review

[J5] C. V. Beojone and N. Geroliminis, "Relocation incentives for ride-sourcing drivers with path-oriented revenue forecasting based on a markov chain model," *Transportation Research Part C: Emerging Technologies*, vol. Under Review, 2023

[J4] C. V. Beojone and N. Geroliminis, "A dynamic multi-region mfd model for ride-sourcing systems with ridesplitting," *Transportation Research Part B: Methodological*, vol. Under Review, 2023. [Online]. Available: `https://arxiv.org/abs/2211.14560`

Journal Articles in Preparation

[J6] C. V. Beojone, Y. Yin, and N. Geroliminis, "Could occupation-based non-linear pricing make ride-pooling a viable alternative during peak hours and improve social-welfare?" *Working Paper (submission planed for September 2023)*, 2023

[J7] C. V. Beojone, P. Zhu, I. İ. Sırmatel, and N. Geroliminis, "A hierarchical control framework for vehicle repositioning in ride-hailing systems," *Working Paper (Full paper submitted to the 25th International Symposium on Transportation and Traffic Theory (ISTTT25))*, 2023

Refereed Conference Articles

[C12] C. V. Beojone and N. Geroliminis, "Providing a revenue-forecasting information scheme as an incentive to relocate compliant ride-sourcing drivers," in *Transportation Research Board 102nd Annual Meeting*, 2023

[C11] C. V. Beojone, N. Geroliminis, and Y. Yin, "Repositioning ridesplitting vehicles through pricing: A two-region simulated study," in *The 11th Triennial Symposium on Transportation Analysis (XI TRISTAN)*, 2022. [Online]. Available: `https://tristan2022.org/Papers/TRISTAN_2022_paper_8692.pdf`

[C10] C. V. Beojone and N. Geroliminis, "A dynamic multi-region mfd model for ride-sourcing systems with ridesplitting," in *Transportation Research Board 101st Annual Meeting*, 2022

[C9] C. V. Beojone and N. Geroliminis, "Macroscopic ride-sourcing model for fleet rebalancing," in *Transportation Research Board 100th Annual Meeting*, 2021

[C8] C. V. Beojone and N. Geroliminis, "Towards sustainable ride-sourcing services: A simulation study on the effects of congestion, fleet size and willingness to share," in *Transportation Research Board 99th Annual Meeting*, 2020

[C7] C. V. Beojone and R. M. de Souza, "A nonstationary hypercube model considering preemptive end-of-shift discipline," in *XIX Latin Ibero-American Conference on Operations Research*, 2018. [Online]. Available: `https://easychair.org/smart-slide/slide/ZNlG#`

[C6] C. V. Beojone, H. H. dos Santos, L. Y. Maruyama, and E. B. Mariano, "Is life-expectancy enough for the hdi?" In *XXIII Symposium of Production Engineering (XXIII SIMPEP)*, 2016. [Online]. Available: `https://www.simpep.feb.unesp.br/abrir_arquivo_pdf.php?tipo=artigo&evento=11&art=315&cad=20785&opcao=com_id`

[C5] C. V. Beojone and R. M. de Souza, "Server scheduling on samu/bauru," in *XXIII Symposium of Production Engineering (XXIII SIMPEP)*, 2016. [Online]. Available: `https://www.simpep.feb.unesp.br/abrir_arquivo_pdf.php?tipo=artigo&evento=11&art=23&cad=4671&opcao=com_id`

[C4] C. V. Beojone and R. M. de Souza, "Randomness and state aggregation on hypercube queueing model," in *XLVIII Brazilian Symposium on Operations Research (XLVIII SBPO)*, 2016, pp. 2183–2195. [Online]. Available: `http://www.din.uem.br/sbpo/sbpo2016/pdf/156160.pdf`

[C3] C. V. Beojone and R. M. de Souza, "State aggregation for hypercube model," in *XXII Symposium of Production Engineering (XXII SIMPEP)*, 2015. [Online]. Available: https://www.simpep.feb.unesp.br/abrir_arquivo_pdf.php?tipo=artigo&evento=10&art=656&cad=4671&opcao=com_id

[C2] C. V. Beojone and R. M. de Souza, "Partial backup on call of the classes in hypercube model with priority in the queue," in *XLVII Brazilian Symposium on Operations Research (XLVII SBPO)*, 2015, pp. 2187–2199. [Online]. Available: http://www.din.uem.br/sbpo/sbpo2015/pdf/141893.pdf

[C1] C. V. Beojone and R. M. de Souza, "Analysis of regionalization samu/bauru using queuing theory," in *XXI Symposium of Production Engineering (XXI SIMPEP)*, 2014. [Online]. Available: https://www.simpep.feb.unesp.br/abrir_arquivo_pdf.php?tipo=artigo&evento=9&art=953&cad=4671&opcao=com_id

Extended Abstracts

[E9] C. V. Beojone and N. Geroliminis, "Providing a revenue-forecasting scheme to relocate groups of ride-sourcing drivers," in *11th Symposium of the European Association for Research in Transportation*, Zurick, Switzerland, 2023

[E8] C. V. Beojone and N. Geroliminis, "Using revenue forecasting to reposition vehicles in ride-sourcing," in *23rd Swiss Transport Research Conference*, Ascona, Switzerland, 2023

[E7] C. V. Beojone and N. Geroliminis, "Ride-sourcing in the urban environment: Traffic interactions, repositioning, and drivers' decisions," in *2nd In-person meeting of Dit4Tram*, Glyphada, Greece, 2022

[E6] C. V. Beojone and N. Geroliminis, "An optimized driver repositioning strategy in ridesplitting with earning estimates: A two-layer dynamic model and control," in *10th Symposium of the European Association for Research in Transportation*, Leuven, Belgium, 2022. [Online]. Available: https://transp-or.epfl.ch/heart/2022/abstracts/235.pdf

[E5] C. V. Beojone and N. Geroliminis, "Repositioning idle vehicles in ridesplitting operations using pricing," in *21st Swiss Transport Research Conference*, Ascona, Switzerland, 2021

[E4] C. V. Beojone and N. Geroliminis, "A path to take passengers from single to shared rides: A study on ridesplitting," in *9th Symposium of the European Association for Research in Transportation*, Lyon, France, 2020. [Online]. Available: https://transp-or.epfl.ch/heart/2020/abstracts/HEART_2020_paper_158.pdf

[E3] C. V. Beojone and N. Geroliminis, "Macroscopic modeling of ride-sourcing services," in *20th Swiss Transport Research Conference*, Ascona, Switzerland, 2020

[E2] C. V. Beojone and N. Geroliminis, "Accessing shared mobility impacts on urban traffic networks through discrete event simulations," in *8th Symposium of the European Association for Research in Transportation*, Budapest, Hungary, 2019. [Online]. Available: https://transp-or.epfl.ch/heart/2019/abstracts/hEART_2019_paper_194.pdf

[E1] C. V. Beojone and N. Geroliminis, "Estimating distances, passenger-vehicle matching and positioning for ridesourcing systems," in *19th Swiss Transport Research Conference*, Ascona, Switzerland, 2019. [Online]. Available: `http://www.strc.ch/2019/Beojone_Geroliminis.pdf`

Training Courses

04.2021 **Scientific Writing Workshop: From Blank Page to Manuscript Draft**
École Polytechnique Fédérale de Lausanne (EPFL)
Instructor: Dr. Kaycie Butler

Research Projects

2021–2023 **Dit4TraM - Distributed Intelligence & Technology for Traffic & Mobility Management**
Sponsor: European Union's Horizon 2020 research and innovation programme (Grant agreement ID: 953783)

2020 **CAROLINA - Real-time pollution City mAp thRough cOLlaborative sensIng aNd Analysis Collaborative and dynamic urban air pollution monitoring**
Sponsor: European Institute of Innovation and Technology (EIT) Urban Mobility

2015–2017 **Studying and Modeling emergency service logistics problems**
Sponsor: São Paulo Research Foundation (grant 2015/05324-4)

2016 **Integrating Data Envelopment Analysis (DEA) and the Hypercube Model for evaluation and comparison of alternative scenarios of operation of Emergency Medical Services**
Sponsor: São Paulo Research Foundation (grant 2014/19051-7)

2014–2015 **Application of the hypercube model on SAMU/Bauru**
Sponsor: São Paulo Research Foundation (grant 2013/02310-7)

Academic Services

Reviewer (Journals)

IEEE Transactions on Intelligent Transportation Systems

Computers & Operations Research

Transportation Research Part C: Emerging Technologies

Journal of Civil Engineering and Management

Transportation Research Part E: Logistics and Transportation Review

Transportation

Omega

Journal of Transport Geography

Economics of Transportation

Reviewer (Conferences)

Transportation Research Board Annual Meeting

Symposium of the European Association for Research in Transportation

Honors and Awards

2015 Ranked 2nd at graduation — *B.Sc., production engineering*, *São Paulo State University (UNESP)*

2021 Mobility Award — *Scholarship for an academic visit to an external research institution*, *Doctoral Program in Civil and Environmental Engineering (EDCE)*

Skills

language Portuguese, English, French (A2)

software MATLAB, Mathematica, Python, LaTeX