Dynamic Control and Modelling of Ride-Sourcing Systems in Large Urban Cities

AMIR HOSEIN VALADKHANI

Master Degree Dissertation

THE UNIVERSITY OF SYDNEY

Supervisor: Dr. Mohsen Ramezani

A thesis submitted in fulfilment of
the requirements for the degree of
Master of Philosophy

School of Civil Engineering
The University of Sydney
Australia

21 September 2021

Contents

List of Figures

CHAPTER 1

Introduction

1.1 Research Background

Increasing the penetration rate of on-demand ride services, including the ride-sourcing systems initiates researches to utilise this technology while minimising its negative side effect on traffic networks or increasing the ride-sourcing companies profit. In this research, we focus on developing a holistic solution for ride-sourcing systems to reduce delay. The solution includes three main components: vehicle-passenger matching, modelling, repositioning of idle vehicles. In this section, we build the historical foundations of ride-sourcing systems that originated from taxi systems. In Sections 2.1 and 3.1, we address the most recent literature that is specifically related to the contributions of each chapter. The contributions of this thesis not only is applicable to improve the level of service of ride-sourcing systems but also can provide better intuition about other point-to-point services like ride-sharing, car pooling, one-way and two-way rebalancing vehicles in carsharing, etc

In this section, we build the historical foundations of ride-sourcing systems that is the backbone of major studies. In Sections 2.1 and 3.1, we address the most recent literature that is specifically related to the contributions of each chapter. In Section 2.1, we focus on the literature of ride-sourcing modeling and vehicle-passenger matching method. Section 3.1 is devoted to the literature of repositioning/re balancing idle vehicles in a ride-sourcing systems.

The foundation of studying ride-sourcing systems originated from taxi and point-to-point systems. The contributions of this thesis not only is applicable to improve the level of

service of ride-sourcing systems but also can provide better intuition about other point-to-point services like ride-sharing, car pooling, one-way and two-way rebalancing vehicle in carsharing, etc. Specific research backgrounds for vehicle-passenger matching method and ride-sourcing modeling are presented in Chapter 2. Moreover, the literature of idle vehicle relocation in ride-sourcing systems is addressed in Chapter 3.

Modeling and control of ride-sourcing systems are interdisciplinary fields of study between the economists, data scientists, and transport engineers. Study of ride-sourcing systems has roots in taxi markets and point-to-point services. From the beginning of 1970, taxi regulation has been studied in many articles. Regulation, granting of franchises or limiting the active taxi numbers are studied in (Beesley 1973) for the first time. (Orr 1969b; Douglas 1972a; De Vany 1975) study the restriction(s) on supply of taxis and controlling the fares. The authors of (Orr 1969b) reveal that the equilibrium of the the regulated taxi market cannot adequately be analyzed by a simple competitive cost-demand theory that is described by (Friedman 1962).

In (Douglas 1972a), cruising taxi system is studied as a market. This article concludes that the differential impacts of regulatory price strategy can be conveniently observed in terms of the effect on the cost of waiting or delay. The results of the article (De Vany 1975) show that the Averch-Johnson (A-J) model as a variant of the usual static simple monopoly model , (Averch and Johnson 1962), and the Chamberlin model as a model explains the short and long run equilibriums that occur under monopolistic competition, , (Chamberlin 1949), are failed for considering the value of excess capacity of consumers in taxi systems. (Shreiber 1975; Coffman and Shreiber 1977) conclude that supplying an appealing model for determination of price and waiting time in unregulated taxicab markets produces a large number of cabs, a short waiting time for customers, and high fares. However, the authors in (Williams 1980) challenge the conclusion of the (Shreiber 1975; Coffman and Shreiber 1977).

In literature of the Transport Engineering, the initial attempts in modelling the cruising of taxis in network level is presented in (Yang and Wong 1998). This article proposes a network model to describe how vacant and occupied taxis search for waiting passengers. The model determine a number of system performance measures at equilibrium such as vacant taxi movements and taxi utilization for a given road network and the customer demand pattern.

This article assumed that each driver tries to minimize individual expected search time before boarding the next customer. In (Xu et al. 1999), a neural network model is developed based on taxi services that are observed in urban area of Hong Kong to model the complex demand-supply relationship in urban taxi services. Deriving the nonlinear macroscopic model of passenger demand, taxi utilization and level-of-services based on a taxi service situation found in the urban area of Hong Kong is studied in (Yang et al. 2000). They assume that (i) taxi drivers move to the customer's destination by the shortest route and (ii) taxi drivers after serving the customer (i.e. reach to the destination) cruise in the network to minimize their expected search time to meet the next customer. The probability that a vacant taxi meets a customer at a particular zone is specified by a logit model. The model is useful to obtain regulatory information to assist with the decisions concerning the restriction over the number of taxi licenses and the fixing of the taxi fare structure.

(Wong et al. 2001) incorporates the congestion effects and customer demand elasticity in the model derived by (Yang et al. 2000). In (Wong et al. 2002), a sensitivity-based solution is proposed for solving the equations of the model suggested by (Wong et al. 2001). The model that is suggested in (Wong et al. 2001) is the advanced form of the simple network model of urban taxi services that is published in (Yang and Wong 1998). A split model is introduced in (Yang et al. 2002b) for prediction of taxi person and vehicle origin-destination matrices. A survey article is published in the beginning of 21th century to review network equilibrium based models and solution algorithms for urban taxi services (Wong et al. 2002).

In (Yang et al. 2002a), a network model is utilized to describe the demand and supply of equilibrium of taxi services under a fare structure and fleet size regulation in an either competitive or monopoly market. The authors of (Yang et al. 2002a) considers spatial structure of the traffic network and customer origin-destination demand pattern. An equilibrium based model to characterize the bilateral searching and meeting between customers and taxis on road networks is studied in (Yang et al. 2010a). This article assumes that the bilateral taxi-customer searching and meeting occurs anywhere in residential and commercial zones or at prescribed taxi stands such as an airport or a railway stations. In (Yang and Yang 2011) the authors introduce a meeting function to characterize the search frictions between vacant

taxis and unserved customers in an aggregated taxi market. A cell-based model to predict local customer-search movements of vacant taxi drivers is proposed in (Wong et al. 2014). It incorporates the modeling principles of the logit-based search model and the intervening opportunity model

In (Ramezani and Nourinejad 2018a), the authors successfully considers the interrelated effects of taxis and normal vehicles (i.e. the influence of taxis on congestion and reversely). The article used macroscopic fundamental diagram (MFD) as the basis of large-scale urban traffic modelling. They develop a dynamic bimodal (cars and taxis) traffic modelling and control strategy (i.e. taxi dispatching) to improve the urban mobility and to mitigate the traffic congestion. The proposed model incorporates two aggregate models, (i) a taxi-passenger meeting function that considers network conditions (i.e. average speed) and (ii) MFD-based traffic dynamics of personal vehicles and taxis. The paper build the bases of how to incorporate matching function with mass conservative equations and how to use MFD for controlling the dispatched taxis.

Another approach in litterateur of the point-to-point transport is utilizing the graph theory to solve the problems. (Santi et al. 2014) propose a taxi sharing system to balance the number of the utilized shared taxis (as benefit indicator of taxi sharing system) and the travel time for the passengers (as discomfort indicator of the passengers). The paper translates the problem into finding the minimum number of the path that covers the shareability network via graph theory. The vertices are trip information (i.e. locations of origin-destination, start and destination times). Finding the minimum number of the taxi cabs (i.e. number of the fleet) to serve all the given trips (specified by origin, destination and start time) without incurring any delay to the passengers is studied in (Vazifeh et al. 2018). (Zhan et al. 2016) develops a taxi service system (i.e. system-wide recommendation sys- tem) for minimizing the cost of empty trips and minimizing the total number of the required taxi to satisfy all the observed trips. Moreover, the paper studies how far is the performance gap between current system and the optimum system by assuming the optimal performance of taxi service system can be measured quantitatively.

The competition between different taxi drivers and different taxi fleets can be model in a game theory framework (Gan et al. 2013; Salhab et al. 2017; Amar and Basir 2018). The agent based pricing for considering the congestion during the peak hour is presented in (Gan et al. 2013). The authors of (Salhab et al. 2017) show how each driver can compute a best-response strategy to the anticipated behavior of the other drivers. This approach results in an approximate Nash equilibrium when the number of the players are limited or not. Cooperative territory allocation approach is studied in (Amar and Basir 2018). This study assumes that negotiation between taxi service providers can reduce the conflict between taxi drivers.

To tackle the complexity of taxi systems, some articles in literature focus on utilizing the artificial intelligence and data-driven approaches to study taxi systems. A real-time recommender system for taxi drivers is suggested in (Huang et al. 2017) to find a next passenger and to start a new trip. This paper uses historical GPS trajectories of taxis. In this article, prediction task is translated to a multi-classification problem rather than regression problem. (Lu et al. 2018) obtains a predictive regional model based on passengers' waiting time. This article employs recurrent neural network (RNN) and deep learning algorithms. (Wang et al. 2017a) investigates the problem of predicting the real-time car-hailing supply-demand which is one of the most important component of an effective scheduling system. The objective is predicting the gap between the car-hailing supply and demand in a certain area in the next few minutes.

Although there are some similarities between taxi systems and ride-sourcing systems that make surveying the taxi systems useful, there are some operational differences between them that make the ride-sourcing systems a disruptive mode. The main differences of ride-sourcing systems are (i) passengers use a third-party digital platform to request a ride, (ii) drivers are dispatched to the passengers by a centralised system, (iii) the real-time locations of the vehicles and passengers are known, (iv) ride-sourcing vehicles are not permitted for street hailing, (v) drivers have flexible work schedules, and (vi) ride-sourcing drivers are independent contractors. These differences encourage researchers to study ride-sourcing system as a new mode of the network. There are many articles in modeling, control, and positioning of taxi systems. In this study we focus on *non-equilibrium* analysis and synthesis of ride-sourcing systems to capture the dynamic of the system for vehicle-passenger matching

and idle vehicle repositioning. Our holistic solution enable us to have sound solution under high variation of demand and supply. The implementation of this method in real-world needs deep understanding of foundation of Macroscopic Fundamental Diagrams and Control Theory.

1.2 Thesis Objective

The goal of the thesis is to develop a holistic framework to improve the service quality of a ride-sourcing system in terms of reducing waiting and unassigned time of the ride-sourcing vehicles and reducing waiting and unassigned time the passengers. To this end, the research approach deals with three main intertwine problems: designing a vehicle-passenger matching method, modeling of a ride-sourcing system in a macroscopic level, and designing a proactive controller for repositioning of idle vehicles.

The specific objectives of this thesis are (i) designing an adaptive spatio-temporal matching method to dynamically find optimum values for vehicle-passenger maximum matching distance and the frequency of the matchings, (ii) developing a validated macroscopic model with capabilities of considering vehicle-passenger matching method and repositioning of idle vehicles to predict the evolution of the state of ride-sourcing system, and (iii) designing a Nonlinear Model Predictive Controller (NMPC) for repositioning of the idle vehicles proactively to the locations with higher probability of being matched to the waiting passengers. The detailed objectives according to the structure of the chapters are:

Part I: Matching Design and Macroscopic Modelling

The key objectives of Chapter 2 are designing a matching method to assign vehicles to waiting passengers and developing a model for formulating a a ride-sourcing system. The matching method consider the join effects of matching frequency and maximum matching distance and level of congestion in dispatching vehicles to passengers. The macroscopic modeling predicts the evolution of the different status of ride-sourcing vehicles (i.e. idle, transferred, dispatched, and occupied) and passengers (i.e. waiting,

assigned, and on-board) in each region of the network. We validate the macroscopic model via microsimulation.

Part II: Idle Vehicle Dynamic Repositioning

The objective of Chapter 3 is developing a dynamic method for relocating idle vehicles proactively to balance the demand of waiting passengers and the supply of ride-sourcing vehicles. In this Chapter, we assume that the drivers obey the ride-sourcing system transfer commands. To this end, we design a NMPC to predict the future evolution of the different status of ride-sourcing vehicles and passengers to minimize the predicted total waiting time of vehicles and total predicted time of passengers over the limited prediction horizon. By microsimulation experiments, we show that the pro-activeness in relocating vehicles reduces the unassigned and waiting time of *vehicles* and *passengers* simultaneously, avoids wild goose chase. Furthermore, we show that the transferring method can be an incentive for majority of vehicles' drivers.

1.3 Thesis Outline

In this thesis we propose a holistic framework that consist of three main subsystem: macroscopic modeling, vehicle-passenger matching method, and transferring of idle vehicles. Chapter 2 focuses on designing a spatio-temporal matching method and macroscopic modelling of a ride-sourcing system. In this chapter, we elaborate different status of ride-sourcing vehicles and passengers in the ride-sourcing system. Also, we illustrate how the proposed macroscopic model comprises dispatching (i.e. matching) and transferring (i.e. redistributor). Furthermore, assessing the performance of the proposed adaptive-spatio temporal matching method and illustrating the accuracy of the model using microsimulation experiments are studied in this chapter. Chapter 2 is organized as an article submitted for publication as:

> *M. Ramezani, AH. Valadkhani, "Dynamic Ride-Sourcing Systems for City-Scale Networks - Part I: Matching Design and Model Formulation and Validation," under review in Transportation Research Part C.*

In Chapter 3, we explain the framework of the dynamic vehicle repositioning system and introduce the mathematical formulation of the transferring system in a state-space form. In the following of this chapter, a discussion on estimation of the prediction of endogenous inputs is presented to increase the controllability of the macroscopic model. We propose a NMPC based on the developed model and estimation of the prediction of endogenous inputs for repositioning idle vehicles proactively. We evaluate the performance of the NMPC in comparison with a reactive controller and with no controller with the developed microsimulation. This chapter is organized as an article submitted for publication as:

> *AH. Valadkhani, M. Ramezani, Dynamic Ride-Sourcing Systems for City-Scale Networks - Part II: Proactive Vehicle Repositioning," under review in Transportation Research Part C.*

In Chapter 4, the thesis is concluded and the future work is presented.

CHAPTER 2

Matching Design and Macroscopic Modelling

The ubiquity of smart-devices enables the foundation for emerging fast-growing ride-sourcing companies that challenges the traditional taxi services. Two design aspects of ride-sourcing systems are: (i) matching mechanism between idle ride-sourcing vehicles and passenger travel requests (i.e. vehicle-passenger matching) and (ii) repositioning mechanism of idle vehicles. In this chapter, we propose a macroscopic non-equilibrium dynamic model of ride-sourcing systems with capabilities of investigating the efficiency of vehicle-passenger matching and idle vehicle repositioning method. A spatio-temporal vehicle-passenger matching method is introduced to determine dynamically and jointly the matching time instances and maximum matching distances to minimize passengers' waiting time (i.e. from the travel request until the pickup) while considering the level of congestion of the network. Designing a controller for repositioning idle vehicles to balance vehicle supply and passenger travel demand based on the proposed model is scrutinized in the companion chapter (Chapter 3). The accuracy of the proposed model and the performance of the matching method under noticeable variation of traffic congestion and passenger travel requests are investigated with microsimulation. The microsimulation results demonstrate accuracy of the model in predicting the evolution of the number of the ride-sourcing vehicles in different states (e.g. idle, transferred, dispatched, and occupied) and passengers (e.g. waiting and assigned) in each region of the network. Furthermore, the proposed matching method pinpoints its effectiveness by reducing reserved and delay times of ride-sourcing vehicles and passengers.

2.1 Introduction

Advancements in information and communication technologies result in growing trends in using ride-sourcing services because of their convenience and affordability. Ride-sourcing service providers like Uber, DiDi, Lyft and their competitors provide on-demand point-to-point services for passengers through an online platform using a fleet of vehicles owned by self-scheduled drivers. The platform requires the service provider to implement strategies for vacant vehicle-passenger matching and transferring of idle vehicles to the travel demand hot-spots to increase profit while ensuring a satisfactory level-of-service such as keeping the expected passenger waiting time below a threshold. In this chapter, we introduce a dynamic non-equilibrium macro-model, as a set of first-order differential equations, to represent dynamics of a ride-sourcing system. Moreover, we develop a method for vehicle-passenger matchings to obtain optimum values for matching time instances (i.e., the time instance when the vehicle-passenger matching method is executed) and the maximum matching distance (i.e., discarding vehicle-passenger matches whose distance is longer than the maximum matching distance) to minimize passengers' waiting time. Capabilities of the suggested model for designing a controller as a dynamic spatial distributor of idle ride-sourcing vehicles is discussed in the companion chapter (Chapter 3).

Modeling ride-sourcing systems exhibit traits of modeling cruising taxis. They are studied extensively with the focus of *equilibrium* analysis which reflects the stable steady-state behavior of the system. The initial steps were taken by equilibrium-based macroscopic models in (Manski and Wright 1967; Orr 1969a; Douglas 1972b). Later, in (Yang and Wong 1998), a stationary model at the equilibrium point is established to formulate movements of cruising taxis. This model is further developed by considering the effect of congestion and passenger demand elasticity in (Wong et al. 2001). The steady-state effect of bilateral taxi-passenger searching and meeting behavior at the equilibrium point of cruising taxi systems is studied in (Wong et al. 2005; Yang et al. 2010a). In (Wong et al. 2015), a two-stage equilibrium-based model is proposed to predict zonal and circulating movements of cruising taxis. In (Yang et al. 2010b; Yang and Yang 2011), the effect of search friction between vacant taxis and passengers on the equilibrium of the taxi system is scrutinized.

A Ride-sourcing market is studied in (Zha et al. 2016) by proposing an equilibrium-based macroscopic model to capture the taxi-passenger meeting dynamics with external matching function. An equilibrium model in a hybrid market with the coexistence of cruising taxi and ride-sourcing systems is investigated in (Qian and Ukkusuri 2017). (Ramezani and Nourinejad 2018a) initiates non-equilibrium based modeling of cruising taxi systems and proposes a predictive controller to relocate the vacant taxis. A non-equilibrium model for a ride-sourcing market with predictive controller is proposed to maximize the overall profit in (Nourinejad and Ramezani 2019).

One dominant factor for quality of service of ride-sourcing systems is vehicle-passenger matching method. The urgency to improve the matching method requires further investigations after imposing harshest-ever set of regulations for ride-sourcing companies by NY city. This regulation necessitates the ride-sourcing companies to carry a passenger at least 69% of the operating time in Manhattan below 96th St from August 2020. Among the recent studies that address vehicle-passenger matching, (Santi et al. 2014; Vazifeh et al. 2018) focuses on high-capacity vehicle-passenger matching, and (Zhan et al. 2016) uses bipartite matching to evaluate the efficiency of taxi systems in terms of optimal matching and trip integration. A driver-rider matching scheme in a ride-sharing system is proposed in (Wang et al. 2017b) to minimize total system-wide vehicle miles. (Zha et al. 2018b) proposes an aggregate matching model to facilitate the analysis of market equilibrium and spatial pricing. Bipartite graph approach is utilised in (Dandl et al. 2019) to propose a static strategy to jointly matched autonomous vehicles (AVs) to user requests and repositions AVs. In (Hörl et al. 2019), a large-scale agent-based simulation of vehicle fleet management for dispatching and repositioning is studied. A data-driven dispatching method for autonomous taxi systems is proposed in (Hu and Dong 2020) to reduce empty travel distance of autonomous taxis.

The issue of vehicle-passenger matching with unlimited radius has been both empirically observed and theoretically analyzed in (Castillo et al. 2017; Xu et al. 2020; Zhang et al. 2019). The instantaneous joint effect of matching radius and matching interval in vehicle-passenger matching is studied in (Yang et al. 2020) in a static way. In these articles, the matching interval is considered fix and known or is obtained based on the instantaneous number of the idle

vehicles and waiting passengers. This neglects the intertwined effects between predicted value of the matching interval (i.e. successive matching instances), instantaneous spatial distribution of idle vehicles and waiting passengers, and predicted value of average vehicle-passenger matching distance/time. To get more insight into the literature of shared mobility, interested readers can further refer to (Salanova et al. 2011; Agatz et al. 2012; Ho et al. 2018; Wang and Yang 2019).

Equilibrium-based modeling of ride-sourcing systems which is common in literature, considers only the steady-state behavior of the system and does not consider the temporal variation of the systems' states (e.g. the number of the vehicles and passengers). Furthermore, there are challenges about existence, uniqueness, and finite-time reachability of equilibrium points of nonlinear dynamic ride-sourcing models. In this study, we propose a macroscopic non-equilibrium model with potential of considering dynamic vehicle-passenger matching and repositioning of idle vehicles to dynamically track different states of passengers and vehicles in a ride-sourcing system. The model is built upon the augmented Cobb-Douglas matching function and Macroscopic Fundamental Diagrams (MFDs). Impatient passengers and drivers are included in the model by defining stochastic cancellation threshold. Furthermore, to optimize expected passengers' waiting time, a vehicle-passenger matching method is proposed to *dynamically* determine (i) the next matching time instance at which the vehicles will be matched to passengers (enabling time-varying matching frequency) and (ii) the maximum matching distance to discard long-distance matchings (avoiding wild goose chase problem and letting vehicles to remain idle for prospect of transferring). To fill the gap in designing matching method, the proposed method is developed to consider the level of congestion of the network, anticipation of the future boarding time, and joint effect of optimum matching interval and maximum matching distance. In Chapter 3, a Nonlinear Model Predictive Controller (NMPC) is designed based on the developed model of this chapter to proactively and dynamically reposition idle vehicles to suppress the spatial imbalance between the idle vehicles (supply) and waiting passengers (demand). The advantages of the matching method and the validity of the model is investigated via a developed microsimulation benchmark.

The remainder of the article is organized as follow. In Section 2.2, we elaborate different states of ride-sourcing vehicles and passengers in the ride-sourcing system. Also, we illustrate how the proposed model comprises dispatching (i.e. matching) and transferring (i.e. redistributor) in Section 2.2. In Section 2.3, the proposed adaptive spatio-temporal matching algorithm for dispatching idle ride-sourcing vehicles to waiting passengers is presented in detail. The proposed dynamic model for tracking of different states of ride-sourcing vehicles and passengers is introduced in Section 2.4. Section 2.5 is devoted to assessing the performance of the proposed adaptive-spatio temporal matching method and illustrating the accuracy of the model using microsimulation experiments. Finally, the article is concluded in Section 2.6.

2.2 Preliminaries

This section defines terminologies used in this article for representing states of the ride-sourcing vehicles and passengers. Furthermore, we depict the schema of the proposed ride-sourcing system to demonstrate different components of the method and the interactions between them.

2.2.1 State Definition

We design a *ride-sourcing system* as a centralized system for dispatching ride-sourcing vehicles to passengers' origin location and transferring the idle vehicles to the locations with a higher possibility of finding passengers. Let us assume an urban network that is partitioned to a few number of regions. We determine four states for the ride-sourcing vehicles (i.e. idle, dispatched, transferred, and occupied) and three states for the passengers (i.e. waiting, assigned, and on-board). The four states of ride-sourcing vehicles are as follow:

(i) *Idle ride-sourcing vehicle* refers to a vacant vehicle which is not assigned to any passenger's travel request. It cruises randomly until receiving a pick-up or transferring command from the ride-sourcing system. The number of idle ride-sourcing vehicles in Region i at time t is denoted as $c_i^{\mathrm{I}}(t)$.

(ii) *Dispatched ride-sourcing vehicle* is a vacant vehicle assigned to a passenger's travel request. It is sent to the passenger's location through the path recommended by the ride-sourcing system. The dispatched vehicles are not allowed to pick up other passengers along the recommended path. The number of dispatched ride-sourcing vehicles in Region i at time t is denoted as $c_i^{\mathrm{D}}(t)$.

(iii) *Transferred ride-sourcing vehicle* is a vacant vehicle sent to a location with an excess of passenger's travel request through the path recommended by the ride-sourcing system. Transferred vehicles are not assigned to any passenger's travel request. They are used to balance the vehicle supply and travel request demand in different regions of the network. They can be assigned to passenger's travel request by the ride-sourcing system before reaching the hot-spot locations. The number of transferred vehicles in Region i at time t is denoted as $c_i^{\mathrm{T}}(t)$.

(iv) *Occupied ride-sourcing vehicle* is a vehicle servicing a passenger. We assume each vehicle services only one passenger or one group of passengers with the same origin and destination. The number of occupied vehicles in Region i at time t is denoted as $c_i^{\mathrm{O}}(t)$.

The three passenger's states are as:

(i) *Waiting passenger* refers to a passenger who has requested a ride but she/he is not assigned to any ride-sourcing vehicle yet. The number of waiting passengers in Region i at time t is denoted as $p_i^{\mathrm{W}}(t)$.

(ii) *Assigned passenger* is a passenger who is not picked up by a vehicle, but a dispatched ride-sourcing vehicle is assigned to her/him. An assigned passenger cannot match with more than one ride-sourcing vehicle at a time. The number of assigned passengers in Region i at time t is denoted as $p_i^{\mathrm{A}}(t)$.

(iii) *On-board passenger* is a passenger picked up by a vehicle but she/he has not reached her/his destination. The number of on-board passengers in Region i at time t is equal to the number of the occupied vehicles in Region i at time t.

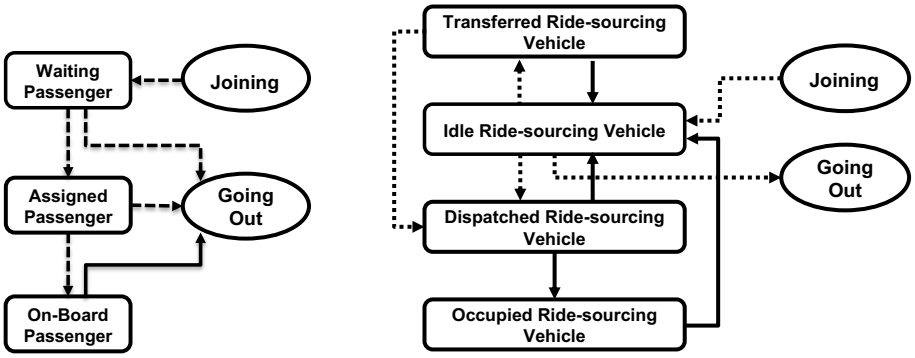
2.2.2 State Transitions

Consider the movements of an *idle ride-sourcing vehicle* in a network which is divided into a number of regions. Once a *waiting passenger* is matched with an idle vehicle, the vehicle becomes *dispatched* and the passenger becomes *assigned*. Then, the dispatched ride-sourcing vehicle starts its travel towards the passenger's pick-up location with the path recommended by the ride-sourcing system. Once, the dispatched vehicle reaches the location of the assigned passenger, the vehicle and passenger become *occupied* and *on-board*, respectively. Once the occupied vehicle reaches the on-board passenger's destination, the state of the vehicle is changed to idle.

An idle ride-sourcing vehicle might be requested to reposition to other region(s) with an excess number of waiting passengers to balance the vehicle sources and passengers' travel demand. If the ride-sourcing system determines such region(s), a number of the idle vehicles become *transferred* and will be guided to moved to that region(s). The transferred vehicles can be assigned to a passenger if a waiting passenger will appear in a proper-distance of the transferred vehicle. Otherwise, the transferred vehicle will reach to the recommended location and becomes idle again.

To reflect the reality more precisely, we assume if an idle ride-sourcing vehicle remains idle for a long period of time, the driver leaves the ride-sourcing system. In addition, if a waiting or an assigned passenger is not picked up for a long time, the passenger cancels the travel request and quit the ride-sourcing service for other travel choices. By canceling a travel request by an assigned passenger, the ride-sourcing vehicle matched to the passenger becomes idle again.

Figure 2.1 illustrates the state diagrams of vehicles and passengers in the ride-sourcing system. The dotted lines emphasize the vehicle's/passenger's state changes instantaneously without a physical trip (e.g. the state of an idle ride-sourcing vehicle becomes transferred once the driver receives the transfer command from the ride-sourcing system.). The solid lines reflect a physical trip in the network (e.g. the dispatched vehicle must reach to the location of the assigned passenger to become occupied.).



(A) (B)

FIGURE 2.1. State diagram of ride-sourcing systems' agents: (a) passengers and (b) ride-sourcing vehicles. Impatient passengers cancel their trips and go out if (i) they remain waiting more than a stochastic threshold or (ii) the assigned vehicle does not pick up the matched passenger before their pickup threshold. The second scenario also affects the dispatched vehicles and change their state to idle.

2.2.3 Proposed Schema

Interconnections between *plant* (i.e. traffic network), *model*, *dispatching subsystem* and *transfer subsystem* are depicted in Figure 2.2. The dispatching subsystem considers idle/transferred ride-sourcing vehicles and waiting passengers as two independent sets that should be matched together to dynamically find the optimal idle/transferred vehicles dispatching in the sense of minimizing the total cruising time of dispatched vehicles considering future states of the system. It includes optimal myopic matching and adaptive spatio-temporal filtering methods. The former minimizes total matching distances with an assumption of maximal matching between the two independent sets, i.e. idle/transferred ride-sourcing vehicles and waiting passengers. The latter method obtains the result of the optimal myopic to minimize the total passengers' waiting time prediction for finding dynamically and jointly: (i) the optimum next matching instance and (ii) the optimum upper bound of matching distance between idle/transferred vehicles and waiting passengers to avoid long-distance matching.

The ride-sourcing model is a set of first order differential equations to represent mass conservation dynamics of the ride-sourcing vehicles and passengers. Transfer subsystem has two

components: *Transferred Ride-sourcing Controller* and *Link-Level Allocation*. The main component of the transfer subsystem is a macroscopic controller that dynamically recommends inter-regional movements of idle ride-sourcing vehicles in the network to balance the vehicles' supply and waiting passengers in the network. We propose a Nonlinear Model Predictive (NMPC) controller as a transfer subsystem for repositioning idle vehicles in Chapter 3 of this study.

Link-Level Allocation translates the aggregated rate of idle vehicles in Region i that is recommended to be transferred to Region j (i.e. output of the transfer controller) to disaggregated (vehicle-level) commands. It (i) selects a subset from the idle vehicles in Region i (based on their current location and the accumulative time that they have remained idle), (ii) assigns a link in Region j to each selected idle vehicles based on the location of the waiting passengers in Region j, and (iii) determines the path with the shortest distance by using Dijkstra's shortest path algorithm for each selected idle vehicle to reach the designated transferred link. The frequency of triggering transfer and dispatched subsystems are independent of each other. More details of Transfer Subsystem is investigated in Chapter 3 of this study.

2.3 Dispatching Subsystem

The dispatching subsystem which includes optimal myopic and adaptive spatio-temporal filtering methods dynamically determines the optimum maximum matching distance and the optimum next matching instance with respect to minimizing passengers waiting time. The dispatching subsystem in each matching instance considers the location of the waiting passengers and idle/transferred ride-sourcing vehicles as well as the aggregated short-term prediction of arrival rate of new passengers and idle/transferred vehicles. In each optimum matching instance that is obtained from the proposed method, firstly, the optimal myopic method is applied and then the optimum maximum matching distance that is obtained from the proposed method is used in each matching instance to filter the long-distance matchings of optimal myopic method. In Subsection 2.3.1, the optimal myopic method is explained that is built upon maximum matching problem of a bipartite graph. It is a classic problem and has

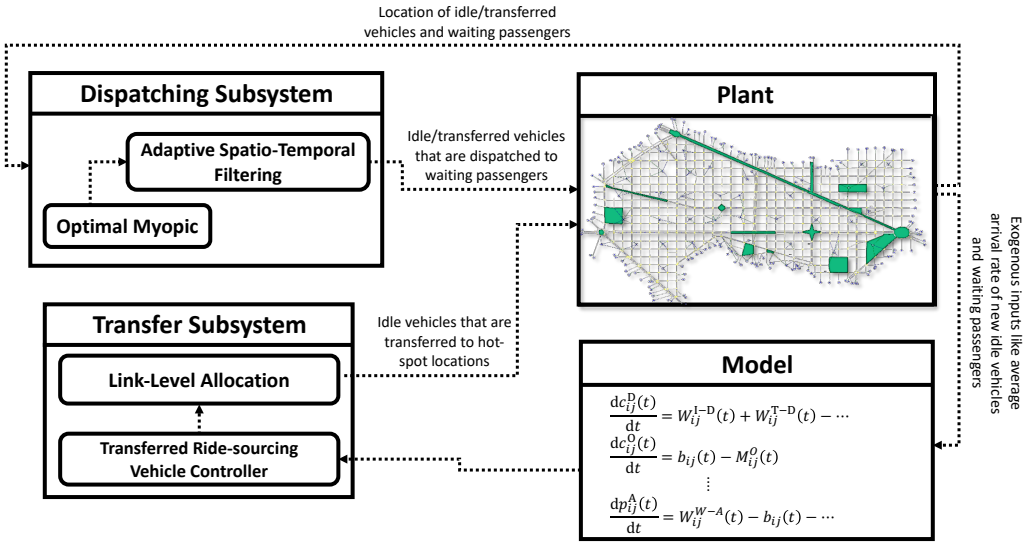


FIGURE 2.2. Schematic of the proposed ride-sourcing system. Dispatching subsystem matches the idle/transferred vehicles to the waiting passengers. The Dispatching Subsystem obtains the required information directly from the plant. The Model is utilised in Transfer Subsystem to predict the effect of idle vehicle repositioning on evolution of each state of the ride-sourcing system. The output of the transfer subsystem is the idle vehicles that are altered to transferred vehicles and are guided to designated locations with excess of passengers.

been widely studied in literature (Agatz et al. 2011; Zhan et al. 2016; Vazifeh et al. 2018). Subsection 2.3.2 introduces the proposed method for finding the optimum matching intervals and maximum matching distances.

2.3.1 Optimal Myopic matching

The optimal matching between idle/transferred ride-sourcing vehicles and waiting passengers at every matching instance is determined by solving the minimum weighted matching problem for a bipartite graph. It minimizes the total matching distances between idle/transferred vehicles and waiting passengers (or equivalently pickup times). We construct the problem as a bipartite graph by considering, (i) V_1 as the set of idle and transferred vehicles, (ii) V_2 as the set of the waiting passengers, and (iii) E as the edges connecting each element of V_1 to V_2. The sets of V_1 and V_2 are disjoint and independent. The weight of each edge, $w(e)$, is the distance between the idle/transferred ride-sourcing vehicle and the waiting passenger (i.e. the

elements of the V_1 and V_2). We obtain the minimum weighted matching for the bipartite graph using integer linear programming method:

$$\begin{aligned} &\text{minimize} \sum_{e \in E} x_e w(e), \\ &\text{s.t.} \quad \sum_{e \sim v} x_e \leq 1 \quad \forall v \in \{V_1 \cup V_2\} \quad \& \quad x_e \in \{0,1\} \qquad \forall e \in E, \end{aligned} \tag{2.1}$$

where, $w(e)$ is the weight of each edge $e \in E$ and $e \sim v$ denotes e is an incident on v. The number of the matching is the minimum of cardinalities of set V_1 and set V_2. Equation 2.1 is a static optimization problem that minimizes total matching distances. It may suffer from matching vehicles with long-distance passengers that wastes their reserved time. Also, it considers only the *current* location of the idle/transferred ride-sourcing vehicles and waiting passengers at each *fixed* matching interval. We consider the effect of future state of the system, i.e. arrival of idle/transferred ride-sourcing vehicles and waiting passengers, congestion of the network, and the joint relationship of matching intervals and discarding long-distance vehicle-passenger matchings in the proposed method in Subsection 2.3.2.

2.3.2 Adaptive Spatio-Temporal Matching Method

The dispatching subsystem includes optimal myopic and adaptive spatio-temporal filtering methods. It dynamically determines the maximum value for the matching distance between idle/transferred ride-sourcing vehicles and waiting passengers as well as occurrence time of the next matching. Changing the instantaneous maximum matching distance has immediate and future effects on vehicle-passenger matchings. Increasing the maximum matching distance immediately increases the average distance of vehicle-passenger matchings and reduce the vehicles' and passengers' unassigned time. Regarding the future effect, increasing the maximum matching distance increases vehicles' and passengers' reserved time. In optimal myopic method, increasing the frequency of the matchings (i.e. decreasing the time between two successive matching instances) causes assigning an idle/transferred vehicle to a waiting passenger once they become available in the network without considering the effect of new passengers and vehicles that might appear in the network relatively just after the matching instance time. Hence, this might lead to matchings that are unreasonably long. In other

words, increasing the matching frequency decreases the *unassigned time* but increases the expected *reserved time*. Unassigned time for a ride-sourcing vehicle (passenger) is the time between the idle (waiting) state and dispatched (assigned) state. The reserved time for a ride-sourcing vehicle (passenger) is the time takes a dispatched vehicle (assigned passenger) becomes occupied (on-board).

The adaptive spatio-temporal matching method aims at minimizing the expected total passengers' waiting time (i.e. unassigned time plus the reserved time) to jointly determine the optimum time of the next matching instance and to discard the vehicle-passenger long-distance matchings. To this end, first, we use the optimal myopic vehicle-passenger matchings (i.e. considering the matching without discarding long-distance matchings) by solving Equation 2.1. Subsequently, we define an expected passengers' waiting time based on the obtained values of the optimal myopic vehicle-passenger matchings. Then, we determine the optimum value of next matching instance and the maximum distance for vehicle-passenger matchings with respect to minimizing the defined expected total passengers' waiting time. The expected total passengers' waiting time consists of four components: (i) estimated total reserved of matchings, (ii) estimated total unassigned time of waiting passengers, (iii) predicted total reserved time for matchings, and (iv) predicted total unassigned time for waiting passengers. Figure 2.3 illustrates the relationship among these components. We assume that the average arrival rate of new waiting passengers and idle vehicles are known.

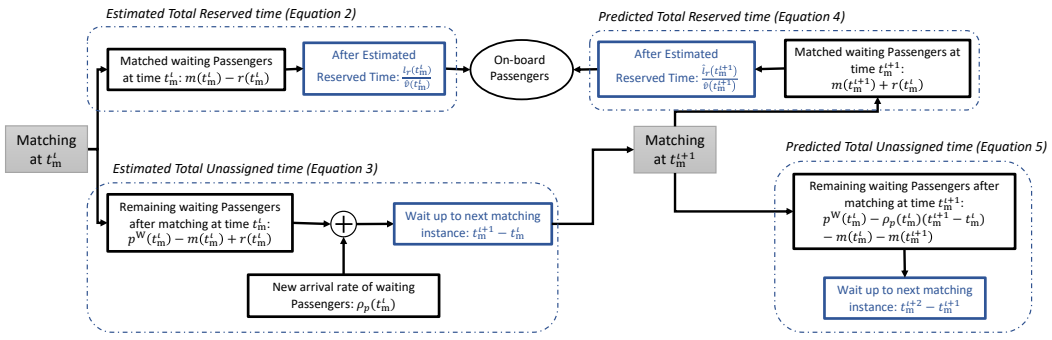


FIGURE 2.3. Schematic of relation between estimated total reserved time, Equation 2.2, estimated total unassigned time, Equation 2.3, predicted total reserved time, Equation 2.4, and predicted total unassigned time, Equation 2.5. Blue boxes represent the time components and black boxes illustrate passengers.

The expected total passengers' waiting time, $\hat{T}_{\mathrm{P}}^{t_{\mathrm{m}}^{\iota+1}-t_{\mathrm{m}}^{\iota}}$, between two successive matching time instances, t_{m}^{ι} and $t_{\mathrm{m}}^{\iota+1}$, is sum of four main parts:

(i) Estimated total reserved time of matchings at t_{m}^{ι} is:

$$\begin{aligned}\hat{T}_{\mathrm{R}}(t_{\mathrm{m}}^{\iota}) &= \frac{\bar{l}_r(t_{\mathrm{m}}^{\iota})}{\hat{v}(t_{\mathrm{m}}^{\iota})}\big(m(t_{\mathrm{m}}^{\iota}) - r(t_{\mathrm{m}}^{\iota})\big) \\ &= \frac{\bar{l}_r(t_{\mathrm{m}}^{\iota})}{\hat{v}(t_{\mathrm{m}}^{\iota})}\Big(\min\big(c^{\mathrm{I}}(t_{\mathrm{m}}^{\iota}) + c^{\mathrm{T}}(t_{\mathrm{m}}^{\iota}), p^{\mathrm{W}}(t_{\mathrm{m}}^{\iota})\big) - r(t_{\mathrm{m}}^{\iota})\Big),\end{aligned} \tag{2.2}$$

where, $\hat{T}_{\mathrm{R}}(t_{\mathrm{m}}^{\iota})$ is the estimated total reserved time for the idle/transferred vehicles assigned to the waiting passengers at t_{m}^{ι}. $\bar{l}_r(t_{\mathrm{m}}^{\iota})$ is the average distance of optimum vehicle-passenger matchings after discarding $r(t_{\mathrm{m}}^{\iota})$ long-distance matchings from the solution of Equation 2.1 at time instance t_{m}^{ι}. The number of the matchings before discarding at time t_{m}^{ι} is denoted by $m(t_{\mathrm{m}}^{\iota})$. The number of the idle vehicles, transfer vehicles, and waiting passengers at time t_{m}^{ι} are indicated by $c^{\mathrm{I}}(t_{\mathrm{m}}^{\iota})$, $c^{\mathrm{T}}(t_{\mathrm{m}}^{\iota})$, and $p^{\mathrm{W}}(t_{\mathrm{m}}^{\iota})$. $\hat{v}(t_{\mathrm{m}}^{\iota})$ denotes the estimated network speed at time instance t_{m}^{ι}.

(ii) Estimated total unassigned time of the waiting passengers remaining in the network after the vehicle-passenger matching at t_{m}^{ι} with discarding $r(t_{\mathrm{m}}^{\iota})$ long-distance matching, $\hat{T}_{\mathrm{D}}(t_{\mathrm{m}}^{\iota})$, is:

$$\hat{T}_{\mathrm{D}}(t_{\mathrm{m}}^{\iota}) = \Big(p^{\mathrm{W}}(t_{\mathrm{m}}^{\iota}) - m(t_{\mathrm{m}}^{\iota}) + r(t_{\mathrm{m}}^{\iota})\Big)\left(t_{\mathrm{m}}^{\iota+1} - t_{\mathrm{m}}^{\iota}\right). \tag{2.3}$$

(iii) Predicted total reserved time for matchings at $t_{\mathrm{m}}^{\iota+1}$, $\hat{T}_{\mathrm{R}}(t_{\mathrm{m}}^{\iota+1})$ is:

$$\begin{aligned}\hat{T}_{\mathrm{R}}(t_{\mathrm{m}}^{\iota+1}) &= \frac{\hat{\bar{l}}_r(t_{\mathrm{m}}^{\iota+1})}{\hat{v}(t_{\mathrm{m}}^{\iota+1})}\big(m(t_{\mathrm{m}}^{\iota+1}) + r(t_{\mathrm{m}}^{\iota})\big) \\ &= \frac{\hat{\bar{l}}_r(t_{\mathrm{m}}^{\iota+1})}{\hat{v}(t_{\mathrm{m}}^{\iota})}\Bigg(\min\Big(c^{\mathrm{I}}(t_{\mathrm{m}}^{\iota}) + c^{\mathrm{T}}(t_{\mathrm{m}}^{\iota}) + \rho_{\mathrm{c}}(t_{\mathrm{m}}^{\iota})\left(t_{\mathrm{m}}^{\iota+1} - t_{\mathrm{m}}^{\iota}\right) - m(t_{\mathrm{m}}^{\iota}), \\ &\qquad p^{\mathrm{W}}(t_{\mathrm{m}}^{\iota}) + \rho_{\mathrm{p}}(t_{\mathrm{m}}^{\iota})\left(t_{\mathrm{m}}^{\iota+1} - t_{\mathrm{m}}^{\iota}\right) - m(t_{\mathrm{m}}^{\iota})\Big) + r(t_{\mathrm{m}}^{\iota})\Bigg),\end{aligned} \tag{2.4}$$

with

$$\rho_{\mathrm{c}}(t_{\mathrm{m}}^{\iota}) = \rho_{\mathrm{c}}^{\mathrm{en}}(t_{\mathrm{m}}^{\iota}) + \rho_{\mathrm{c}}^{\mathrm{ex}}(t_{\mathrm{m}}^{\iota}),$$

where, $\hat{\bar{l}}_r(t_{\mathrm{m}}^{\iota+1})$ estimates the average matching distance of matched pairs in a bipartite matching program as Equation 1. This parameter is defined for prediction of the next matching instance, $t_{\mathrm{m}}^{\iota+1}$, and does not reflect the observed average of matched distances. Since the exact locations of idle/transferred vehicles and waiting passengers at time $t_{\mathrm{m}}^{\iota+1}$ are not known, we propose a parsimonious function to estimate the average matching distance of the optimum matching that is only a function of the number of vehicles and passengers. Subsection 2.3.3 explains estimation of $\hat{\bar{l}}_r(t_{\mathrm{m}}^{\iota+1})$. $m(t_{\mathrm{m}}^{\iota+1})$ denotes the number of the matchings at time $t_{\mathrm{m}}^{\iota+1}$ without discarding at time t_{m}^{ι}. The rates of arrival of idle/transferred vehicles and waiting passengers during interval $[t_{\mathrm{m}}^{\iota}, t_{\mathrm{m}}^{\iota+1})$ are denoted by $\rho_{\mathrm{c}}(t_{\mathrm{m}}^{\iota})$ and $\rho_{\mathrm{p}}(t_{\mathrm{m}}^{\iota})$, respectively. $\rho_{\mathrm{c}}(t_{\mathrm{m}}^{\iota})$ has endogenous, $\rho_{\mathrm{c}}^{\mathrm{en}}(t_{\mathrm{m}}^{\iota})$, and exogenous parts, $\rho_{\mathrm{c}}^{\mathrm{ex}}(t_{\mathrm{m}}^{\iota})$. $\rho_{\mathrm{c}}^{\mathrm{en}}(t_{\mathrm{m}}^{\iota})$ captures the rate that occupied or dispatched vehicles become idle. $\rho_{\mathrm{c}}^{\mathrm{ex}}(t_{\mathrm{m}}^{\iota})$ captures the rate that idle ride-sourcing vehicles leave the network because they are not assigned to any passenger for a long time and the vehicles leaving from or entering to the network due to their working hours.

(iv) Predicted total unassigned time for waiting passengers remaining in the network after the vehicle-passenger matching at $t_{\mathrm{m}}^{\iota+1}$, $\hat{T}_{\mathrm{D}}(t_{\mathrm{m}}^{\iota+1})$, is:

$$\begin{aligned}\hat{T}_{\mathrm{D}}(t_{\mathrm{m}}^{\iota+1}) = \\ &\Big(p^{\mathrm{W}}(t_{\mathrm{m}}^{\iota}) + \rho_{\mathrm{p}}(t_{\mathrm{m}}^{\iota})\left(t_{\mathrm{m}}^{\iota+1} - t_{\mathrm{m}}^{\iota}\right) - \left(m(t_{\mathrm{m}}^{\iota}) - r(t_{\mathrm{m}}^{\iota})\right) - \left(m(t_{\mathrm{m}}^{\iota+1}) + r(t_{\mathrm{m}}^{\iota})\right)\Big) \\ &\qquad \times \left(t_{\mathrm{m}}^{\iota+2} - t_{\mathrm{m}}^{\iota+1}\right) \\ &= \Big(p^{\mathrm{W}}(t_{\mathrm{m}}^{\iota}) + \rho_{\mathrm{p}}(t_{\mathrm{m}}^{\iota})\left(t_{\mathrm{m}}^{\iota+1} - t_{\mathrm{m}}^{\iota}\right) - m(t_{\mathrm{m}}^{\iota}) - m(t_{\mathrm{m}}^{\iota+1})\Big)\left(t_{\mathrm{m}}^{\iota+2} - t_{\mathrm{m}}^{\iota+1}\right).\end{aligned} \tag{2.5}$$

In Equation 2.5, we assume the discarded matchings at time t_{m}^{ι}, i.e. $r(t_{\mathrm{m}}^{\iota})$, will be matched at time $t_{\mathrm{m}}^{\iota+1}$. We update this assumption in rolling time horizon.

By assuming optimal myopic at time $t_{\mathrm{m}}^{\iota+1}$, the expected total passengers' waiting time is:

$$\hat{T}_{\mathrm{P}}^{t_{\mathrm{m}}^{\iota+1} - t_{\mathrm{m}}^{\iota}} = \hat{T}_{\mathrm{R}}(t_{\mathrm{m}}^{\iota}) + \hat{T}_{\mathrm{D}}(t_{\mathrm{m}}^{\iota}) + \hat{T}_{\mathrm{R}}(t_{\mathrm{m}}^{\iota+1}) + \hat{T}_{\mathrm{D}}(t_{\mathrm{m}}^{\iota+1}). \tag{2.6}$$

We obtain the optimum value for the next matching time, $t_{\mathrm{m}}^{\iota+1}$, and the number of discarded long-distance matchings, $r(t_{\mathrm{m}}^{\iota})$, by minimizing the total passengers' waiting time prediction:

We obtain the optimum value for the next matching time, $t_{\mathrm{m}}^{\iota+1}$, and the number of discarded long-distance matchings, $r(t_{\mathrm{m}}^{\iota})$, by minimizing the total passengers' waiting time prediction as determined in Equation 7. Vehicle-passenger matches with matching distances longer than the maximum matching distance are discarded at each matching time instance.

$$\begin{aligned} &\underset{r(t_{\mathrm{m}}^{\iota}),t_{\mathrm{m}}^{\iota+1}}{\text{minimize}} \ \left(\hat{T}_{\mathrm{P}}^{t_{\mathrm{m}}^{\iota+1}-t_{\mathrm{m}}^{\iota}}\right), \\ &\quad \text{s.t.} \quad 0 < t_{\mathrm{m}}^{\iota} < t_{\mathrm{m}}^{\iota+1} \leq t_{\mathrm{m}}^{\iota+2}, \quad t^{\iota+2}{}_{\mathrm{m}} \leq t^{\iota_{\mathrm{m}}} + t_{\mathrm{m}}^{\max}, \quad 0 \leq r(t_{\mathrm{m}}^{\iota}) \leq m(t_{\mathrm{m}}^{\iota}) \end{aligned} , \tag{2.7}$$

where, $t_{\mathrm{m}}^{\max}$ is a predefined upper bound for the next matching time instance. To practically solve the discrete-continues optimization problem of Equation 7, after solving the optimal myopic matching program, we discretize the time domain, $[t_{\mathrm{m}}^{\iota}, t_{\mathrm{m}}^{\max}]$, and iteratively evaluate Equation 6 for any specific $t_{\mathrm{m}}^{\iota+1}$ and $r(t_{\mathrm{m}}^{\iota})$ that satisfy the criteria of Equation 7 to determine the optimum values. Choosing $t_{\mathrm{m}}^{\max}$ and sampling period for discretization are important factors to reach an attainable solution computationally. The major time complexity of solving the optimization problem is stemmed from solving the optimal myopic matchings that can be relaxed significantly from $O(|V|^4)$ to $O(\sqrt{|V|})$ by adopting an approximate method instead of exact method (Vazirani 1994). To this end, we choose $t_{\mathrm{m}}^{\max} = 120$ [sec] and sampling period of 5 [sec]. The performance of the proposed adaptive spatio-temporal matching method for dispatching idle/transferred vehicles to waiting passengers is investigated in Section 2.5.

2.3.3 Estimating a Macro-Function for Average Optimal Myopic Matching Distance

A closed-form macro-function for estimating average optimal myopic matching distance, $\hat{\bar{l}}$, is needed for predicting the total reserved time, see Equation 2.4. To this end, we propose Algorithm 1 to estimating the average optimal myopic matching distance as a function of the number of the idle/transferred ride-sourcing vehicles and waiting passengers, independent of

their location. Algorithm 1 generates the coordinates of idle/transferred ride-sourcing vehicles and waiting passengers randomly with uniform distribution inside a network. Then, a bipartite graph, $G(V_1, V_2, E)$, is built in which the weights of the edges are the Manhattan distance between idle/transferred ride-sourcing vehicles and waiting passengers. The average optimal matching is determined by solving the matching problem on graph G to minimize the sum of matching weights as in Equation 2.1. To tackle the stochasticity of spatial distribution of ride-sourcing vehicles and passengers, this procedure is repeated N^{itr} times for each number of idle/transferred ride-sourcing vehicles and waiting passengers. This experiment is conducted for a range of passengers, 1 to N^{P}, and a range of vehicles, 1 to N^{T}, where N^{P} and N^{T} denote the maximum numbers of passengers and vehicles that are considered, respectively.

We run the algorithm on a high performance computer (HPC) for N^{itr}=100, N^{P}=50, and N^{T}=50. The result, see Figure 2.4(b), reveals that the variations of average optimum matching distance with respect to the number of waiting passengers and idle/transferred ride-sourcing vehicles are:

$$\begin{cases} \frac{\partial \hat{\bar{l}}}{\partial (c^{\text{I}}+c^{\text{T}})} > 0,\ \frac{\partial \hat{\bar{l}}}{\partial p^{\text{W}}} < 0 & \text{if} \quad c^{\text{I}} + c^{\text{T}} < p^{\text{W}} \\ \frac{\partial \hat{\bar{l}}}{\partial (c^{\text{I}}+c^{\text{T}})} < 0,\ \frac{\partial \hat{\bar{l}}}{\partial p^{\text{W}}} > 0 & \text{if} \quad p^{\text{W}} < c^{\text{I}} + c^{\text{T}} \end{cases} .$$

Accordingly, the following symmetric form is suggested for the average optimum matching distance:

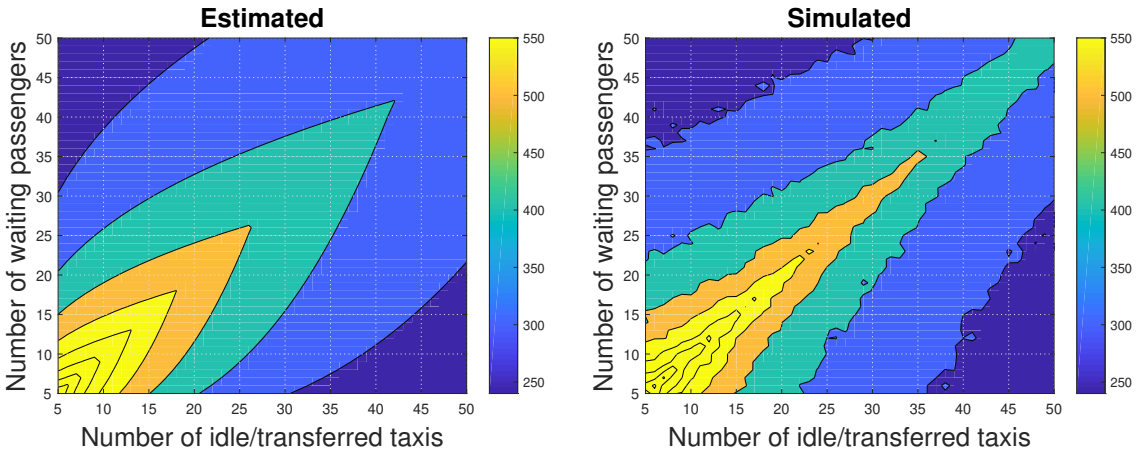
$$\hat{\bar{l}} = \begin{cases} \theta (c^{\text{I}} + c^{\text{T}})^{\zeta_1} p^{\text{W}\zeta_2} & \text{if} \quad c^{\text{I}} + c^{\text{T}} \leq p^{\text{W}} \\ \theta (c^{\text{I}} + c^{\text{T}})^{\zeta_2} p^{\text{W}\zeta_1} & \text{if} \quad p^{\text{W}} < c^{\text{I}} + c^{\text{T}} \end{cases} \tag{2.8}$$

where, $\theta > 0$, $\zeta_1 > 0$, and $\zeta_2 < 0$ are parameters that can be readily estimated using the Least Square method. By utilizing the generated data in algorithm 1, the estimated values of the parameters are: $\hat{\theta} = 2394.57$, $\hat{\zeta_1} = 0.245$, and $\hat{\zeta_2} = -0.724$, where, $R^2 = 0.93$. Figure 2.4 compares the simulated and estimated average optimum matching distances.

Algorithm 1: Pseudocode for estimating the average optimum matching distance

Result: $\{\hat{\bar{l}}_{n^{\mathrm{P}} \times n^{\mathrm{T}}} |\ n^{\mathrm{P}} = 1 : N^{\mathrm{P}},\ n^{\mathrm{T}} = 1 : N^{\mathrm{T}}\}$

Spatial initialization;
for $n^{\mathrm{T}} = 1 : N^{\mathrm{T}}$ **do**
  for $n^{\mathrm{P}} = 1 : N^{\mathrm{P}}$ **do**
    for $k = 1 : N^{\mathrm{itr}}$ **do**
      $\{(x_i^{\mathrm{P}}, y_i^{\mathrm{P}}) |\ i = 1 : n^{\mathrm{P}}\} \leftarrow$ Generate coordinates of n^{P} passengers randomly;
      $\{(x_j^{\mathrm{T}}, y_j^{\mathrm{T}}) |\ j = 1 : n^{\mathrm{T}}\} \leftarrow$ Generate coordinates of n^{T} ride-sourcing vehicles randomly;
      $G(V_1, V_2, E) \leftarrow$ Initialization a bipartite graph with $|V_1| = n^{\mathrm{P}}$ and $|V_2| = n^{\mathrm{T}}$;
      for $i = 1 : n^{\mathrm{P}}$ **do**
        for $j = 1 : n^{\mathrm{T}}$ **do**
          $w_{ij} \leftarrow$ Manhattan distance between passenger i and ride-sourcing vehicle j;
          $d_{ij} \leftarrow$ Add an edge to connect vertices $(v_1^{\mathrm{i}}, v_2^{\mathrm{j}})$ with weight w_{ij};
      $M_k \leftarrow$ Find optimal matching of G to minimize the sum of matching weights;
      $\hat{l}_k \leftarrow$ Find the average of matched weights;
    $\hat{\bar{l}}_{n^{\mathrm{P}} \times n^{\mathrm{T}}} \leftarrow$ Find the average of $\{\hat{l}_k |\ k = 1 : N^{\mathrm{itr}}\}$;



(A) (B)

FIGURE 2.4. (a) Estimated average of optimum matching distance and (b) simulated average of optimum matching distance.

2.4 Model Formulation

Consider a network divided into a set of $\mathbb{R}$ regions. The proposed ride-sourcing model formulates the interaction of waiting passengers, assigned passengers, idle ride-sourcing vehicles, occupied ride-sourcing vehicles, dispatched ride-sourcing vehicles, and transferred ride-sourcing vehicles in different regions as illustrated in the state diagrams of Figure 2.1. We develop our model based on MFD and Cobb-Douglas meeting function.

MFD is a macroscopic relationship between the number of vehicles (accumulation) and aggregated outflow (or weighted flow of links, production) in homogeneous regions. Existence of this relationship is empirically presented in (Geroliminis and Daganzo 2008). The effect of congestion heterogeneity on MFD properties is studied in (Mahmassani et al. 2013; Ramezani et al. 2015). A three-step clustering algorithm is proposed in (Saeedmanesh and Geroliminis 2016) to partition heterogeneous networks into a few connected homogeneous regions. This offers traffic management opportunities based on MFD modeling such as perimeter control (Li et al. 2021; Haddad and Zheng 2020), regional routing (Ingole et al. 2020), and regional pricing (Zheng et al. 2012), among others. Applicability of using MFD on modeling of on-demand mobility services including cruising taxi systems, ride-sharing systems, and pricing

of ride-sourcing systems are studied in (Ramezani and Nourinejad 2018a; Alisoltani et al. 2021; Nourinejad and Ramezani 2019).

In this study, the MFDs are used to estimate the internal and inter-regional flows [1] of occupied and transferred ride-sourcing vehicles as well as the inter-regional flow for the dispatched ride-sourcing vehicles. Cobb-Douglas meeting functions are utilized to model the boarding rate of assigned passengers and the rate that dispatched ride-sourcing vehicles become occupied. The reason for differentiating between estimating the internal flow of the dispatched ride-sourcing vehicles and occupied/transferred ride-sourcing vehicles is their travel behaviour. The average trip length for internal flow of the dispatched ride-sourcing vehicles is highly sensitive to the number of the dispatched ride-sourcing vehicles and assigned passengers. If we assume the number of the dispatched ride-sourcing vehicles is constant, the higher the number of the assigned passengers, the less the average trip length is. Hence, we use Cobb-Douglas meeting function to consider the number of dispatched ride-sourcing vehicles and assigned passengers simultaneously in boarding rate. However, the average trip lengths for the internal and inter-regional flow of the occupied/transferred ride-sourcing vehicles as well as inter-regional flow of the dispatched ride-sourcing vehicles are independent of the number of the passengers. That is, their travel patterns are similar to the normal vehicles. Therefore, we utilize MFDs to estimate their trip completions.

The boarding function between dispatched ride-sourcing vehicles and assigned passengers in Region i, b_i, is defined by a Cobb-Douglas type meeting function to consider the friction and congestion (Yang and Yang 2011; Ramezani and Nourinejad 2018a):

$$\begin{aligned} b_i(t) &= K_i c_{ii}^{\mathrm{D}}(t)^{\alpha_i} p_i^{\mathrm{A}}(t)^{\beta_i} v_i(t)^{\gamma_i} && \forall i \in \mathbb{R}, \\ p_i^{\mathrm{A}}(t) &= \sum_{j \in \{\mathbb{U}_i \cup i\}} p_{ij}^{\mathrm{A}}(t) && \forall i \in \mathbb{R}, \end{aligned} \tag{2.9}$$

where, K_i is the total productivity factor of Region i. The number of the dispatched ride-sourcing vehicles in Region i with pickup location in Region i at time t is denoted by $c_{ii}^{\mathrm{D}}(t)$. The total number of the assigned passengers in Region i and the number of the

[1] In MFD literature this term is typically named as transfer flow. However to avoid confusion with the transfer state of ride-sourcing vehicles, we use the inter-region flow term in this chapter.

assigned passengers in Region i with destination in Region j are denoted by $p_i^{\mathrm{A}}(t)$ and $p_{ij}^{\mathrm{A}}(t)$, respectively. $\mathbb{U}_i$ is the set of regions that are in the direct vicinity of Region i. $v_i(t)$ is the average speed in Region i. Considering the average speed in the boarding function reflects the level of congestion on the meeting rate (Ramezani and Nourinejad 2018a). We show the elasticities with respect to the number of dispatched vehicles, assigned passengers, and average speed by α_i, β_i, and γ_i, respectively. K_i, α_i, β_i, and γ_i are constant parameters that their values can be estimated from the field or simulated data.

By defining the boarding rate in Region i as Equation 2.9, the boarding rate in Region i with the final destination in Region j at time t can be approximated using from the following equation. We assume the boarding rate is proportional to assigned passengers:

$$b_{ij}(t) = \frac{p_{ij}^{\mathrm{A}}(t)}{p_i^{\mathrm{A}}(t)} b_i(t) \qquad \forall i \in \mathbb{R} \text{ and } j \in \{\mathbb{U}_i, i\} \,. \tag{2.10}$$

We utilize the MFD defined by production function to obtain the inter-regional flows of occupied vehicles, $M_{ij}^{\mathrm{O}}(t)$, transferred vehicles, $M_{ij}^{\mathrm{T}}(t)$, and dispatched vehicles, $M_{ij}^{\mathrm{D}}(t)$. We assume the inter-regional flows are proportional to the accumulations:

$$M_{ij}^{\mathrm{O}}(t) = \frac{c_{ij}^{\mathrm{O}}(t)}{n_i(t)} \frac{P_i(n_i(t))}{l_{ij}^{\mathrm{O}}(t)} \qquad \forall i \in \mathbb{R} \text{ and } j \in \mathbb{U}_i \,, \tag{2.11}$$

$$M_{ij}^{\mathrm{T}}(t) = \frac{c_{ij}^{\mathrm{T}}(t)}{n_i(t)} \frac{P_i(n_i(t))}{l_{ij}^{\mathrm{T}}(t)} \qquad \forall i \in \mathbb{R} \text{ and } j \in \mathbb{U}_i \,, \tag{2.12}$$

$$M_{ij}^{\mathrm{D}}(t) = \frac{c_{ij}^{\mathrm{D}}(t)}{n_i(t)} \frac{P_i(n_i(t))}{l_{ij}^{\mathrm{D}}(t)} \qquad \forall i \in \mathbb{R} \text{ and } j \in \mathbb{U}_i \,, \tag{2.13}$$

where, $c_{ij}^{\mathrm{O}}(t)$, $c_{ij}^{\mathrm{T}}(t)$, and $c_{ij}^{\mathrm{D}}(t)$ are respectively the number of the occupied, transferred, and dispatched vehicles in Region i moving to Region j at time t. $n_i(t)$ is the total accumulation of vehicles in Region i including normal vehicles and all types of the ride-sourcing vehicles in Region i. $P_i(n_i(t))$ is the production MFD in Region i that is a function of the total accumulation in Region i. The average trip length for occupied, transferred, and dispatched vehicles in Region i that travel to Region j are denoted by $l_{ij}^{\mathrm{O}}(t)$, $l_{ij}^{\mathrm{T}}(t)$, and $l_{ij}^{\mathrm{D}}(t)$, respectively.

The internal flows for the occupied, $M_{ii}^{\mathrm{O}}(t)$, and the transferred ride-sourcing vehicles, $M_{ii}^{\mathrm{T}}(t)$, are:

$$M_{ii}^{\mathrm{O}}(t) = \frac{c_{ii}^{\mathrm{O}}(t)}{n_i(t)} \frac{P_i(n_i(t))}{l_{ii}^{\mathrm{O}}(t)} \qquad \forall i \in \mathbb{R}, \tag{2.14}$$

$$M_{ii}^{\mathrm{T}}(t) = \frac{c_{ii}^{\mathrm{T}}(t)}{n_i(t)} \frac{P_i(n_i(t))}{l_{ii}^{\mathrm{T}}(t)} \qquad \forall i \in \mathbb{R}, \tag{2.15}$$

where, $c_{ii}^{\mathrm{O}}(t)$ and $c_{ii}^{\mathrm{T}}(t)$ are the number of the occupied and transferred ride-sourcing vehicles in Region i with destination in Region i. $l_{ii}^{\mathrm{O}}(t)$ and $l_{ii}^{\mathrm{T}}(t)$ denote the average trip length of occupied and transferred vehicles in Region i with destination in Region i.

Using the above internal and inter-regional flows, we can derive the dynamics of different states of the system as mass conservation formulations. The conservation of the number of the occupied ride-sourcing vehicles is modeled as:

$$\frac{\mathrm{d}c_{ii}^{\mathrm{O}}(t)}{\mathrm{d}t} = b_{ii}(t) + \sum_{j \in \mathbb{U}_i} M_{ji}^{\mathrm{O}}(t) - M_{ii}^{\mathrm{O}}(t) \qquad \forall i \in \mathbb{R}, \tag{2.16}$$

$$\frac{\mathrm{d}c_{ij}^{\mathrm{O}}(t)}{\mathrm{d}t} = b_{ij}(t) - M_{ij}^{\mathrm{O}}(t) \qquad \forall i \in \mathbb{R} \text{ and } j \in \mathbb{U}_i. \tag{2.17}$$

In RHS of Equations 2.16 and 2.17, $b_{ii}(t)$ and $b_{ij}(t)$ are the number of the dispatched ride-sourcing vehicles that become occupied with destination in regions i and j respectively. $\sum_{j \in \mathbb{U}_i} M_{ji}^{\mathrm{O}}(t)$ reflects the rate that the occupied ride-sourcing vehicles from regions of $\mathbb{U}_i$ pass the regions boundary towards Region i. The rate that the occupied ride-sourcing vehicles complete their trips and become idle is $M_{ii}^{\mathrm{O}}(t)$.

The evolution of the number of dispatched ride-sourcing vehicles over time is:

$$\frac{\mathrm{d}c_{ii}^{\mathrm{D}}(t)}{\mathrm{d}t} = W_{ii}^{\text{I-D}}(t) + W_{ii}^{\text{T-D}}(t) + \sum_{j \in \mathbb{U}_i} M_{ji}^{\mathrm{D}}(t) - b_i(t) - R_{ii}^{\mathrm{D}}(t) \qquad \forall i \in \mathbb{R}, \tag{2.18}$$

$$\frac{\mathrm{d}c_{ij}^{\mathrm{D}}(t)}{\mathrm{d}t} = W_{ij}^{\text{I-D}}(t) + W_{ij}^{\text{T-D}}(t) - M_{ij}^{\mathrm{D}}(t) - R_{ij}^{\mathrm{D}}(t) \qquad \forall i \in \mathbb{R} \text{ and } j \in \mathbb{U}_i, \tag{2.19}$$

where, c_{ii}^{D} and c_{ij}^{D} respectively denote the number of the dispatched ride-sourcing vehicles in Region i with pickup location in regions i and j. $W_{ii}^{\text{I-D}}(t)$ and $W_{ij}^{\text{I-D}}(t)$ are the rate of idle ride-sourcing vehicles in Region i that the matching method of ride-sourcing system

dispatches them to regions i and j at time t, respectively. $W_{ii}^{\text{T-D}}(t)$ and $W_{ij}^{\text{T-D}}(t)$ denote the rate of transferred ride-sourcing vehicles in Region i that the matching method dispatches them to regions i and j at time t. Variables, $W_{ii}^{\text{I-D}}(t)$, $W_{ij}^{\text{I-D}}(t)$, $W_{ii}^{\text{T-D}}(t)$, and $W_{ij}^{\text{T-D}}(t)$ are the solution of the dispatching subsystem (see Figure 2.2) at each matching time instance. $R_{ii}^{\text{D}}(t)$ and $R_{ij}^{\text{D}}(t)$ are the rate of the cancellation of dispatched trips at time t in Region i with final destination (i.e., pickup location) in Regions i and j, respectively. They can represent the cancellation of dispatched trips before and after crossing the boundary. For example, consider a ride-sourcing vehicle that is currently in Region 1 with final dispatched destination in Region 4. If the trip is cancelled at time t before crossing the boundary, it will be considered in $R_{14}^{\text{D}}(t)$. Otherwise, it will be considered in $R_{44}^{\text{D}}(t)$.

The number of the transferred ride-sourcing vehicles at each time is obtained from the following equations. We assume that the transferred ride-sourcing vehicles in Region i complete their trips within Region i:

$$\frac{\mathrm{d}c_{ii}^{\text{T}}(t)}{\mathrm{d}t} = \sum_{j\in\mathbb{U}_i} M_{ji}^{\text{T}}(t) - M_{ii}^{\text{T}}(t) - W_{ii}^{\text{T-D}}(t) \qquad \forall i \in \mathbb{R}, \tag{2.20}$$

$$\frac{\mathrm{d}c_{ij}^{\text{T}}(t)}{\mathrm{d}t} = W_{ij}^{\text{I-T}}(t) - M_{ij}^{\text{T}}(t) - W_{ij}^{\text{T-D}}(t) \qquad \forall i \in \mathbb{R} \text{ and } j \in \mathbb{U}_i, \tag{2.21}$$

where, $W_{ij}^{\text{I-T}}(t)$ is the rate that the idle vehicles in Region i are advised to transfer to Region j by the transfer controller. This is an external manipulating (i.e. control) variable that is the result of the transfer subsystem to balance the number of the idle ride-sourcing vehicles and waiting passengers in the network. We propose a predictive controller in Chapter 3 to dynamically determine the optimum values of $W_{ij}^{\text{I-T}}(t)$.

Total number of the idle ride-sourcing vehicles in Region i is evolved based on the following equation:

$$\begin{aligned}\frac{\mathrm{d}c_i^{\text{I}}(t)}{\mathrm{d}t} =& q_i^{c^+}(t) + M_{ii}^{\text{O}}(t) + M_{ii}^{\text{T}}(t) - q_i^{c^-}(t) - \sum_{j\in\{\mathbb{U}_i,i\}} W_{ij}^{\text{I-D}}(t) \\ & - \sum_{j\in\mathbb{U}_i} W_{ij}^{\text{I-T}}(t) + w_i^{\text{I}}(t) \qquad \forall i \in \mathbb{R},\end{aligned} \tag{2.22}$$

where, $q_i^{c^+}(t)$ is the rate of idle ride-sourcing vehicles exogenously enters to the network at time t. $q_i^{c^-}(t)$ denotes the rate of idle ride-sourcing vehicles leave the network at time t due to not being satisfied by the ride-sourcing level of service or because of finishing their working hours. $M_{ii}^{\mathrm{O}}(t)$ and $M_{ii}^{\mathrm{T}}(t)$ reflect the rate of occupied and transferred ride-sourcing vehicles completing their trips and becoming idle. Total number of the idle ride-sourcing vehicles becoming dispatch and transferred in Region i are considered in $\sum_{j\in\mathbb{U}_i} W_{ij}^{\text{I-D}}(t)$ and $\sum_{j\in\mathbb{U}_i} W_{ij}^{\text{I-T}}(t)$ terms. Some idle ride-sourcing vehicles prefer to stop until they are dispatched to an assigned passenger. $w_i^{\mathrm{I}}(t)$ is a stochastic term that represents the rate of the idle vehicles that enter (or leave) Region i at time t because of their random cruising; this might happen to non-dispatched non-transferred idle vehicles that cruise around the boundary of regions.

Based on Figure 2.1, there are three states for the passengers in the proposed ride-sourcing system: (i) waiting passengers, (ii) assigned passengers, and (iii) on-board passengers. The conservation of the on-board passengers over time is exactly similar to the occupied ride-sourcing vehicles but we need to present the conservation of the waiting and assigned passengers. The dynamics of the waiting passengers is ensured by the following equations:

$$\frac{\mathrm{d}p_{ii}^{\mathrm{W}}(t)}{\mathrm{d}t} = q_{ii}^{p^{W+}}(t) - q_{ii}^{p^{W-}}(t) - W_{ii}^{\text{W-A}}(t) \qquad \forall i \in \mathbb{R}, \tag{2.23}$$

$$\frac{\mathrm{d}p_{ij}^{\mathrm{W}}(t)}{\mathrm{d}t} = q_{ij}^{p^{W+}}(t) - q_{ij}^{p^{W-}}(t) - W_{ij}^{\text{W-A}}(t) \qquad \forall i \in \mathbb{R} \text{ and } j \in \mathbb{U}_i, \tag{2.24}$$

where, $p_{ii}^{\mathrm{W}}(t)$ and $p_{ij}^{\mathrm{W}}(t)$ denote the number of waiting passengers in Region i with destinations in Region i and Region j at time t, respectively. The exogenous rate of waiting passenger demand in Region i going to regions i and j are denoted by $q_{ii}^{p^{W+}}(t)$ and $q_{ij}^{p^{W+}}(t)$, respectively. $q_{ii}^{p^{W-}}(t)$ and $q_{ij}^{p^{W-}}(t)$ are the rate of the waiting passengers in Region i with destination in regions i and j leaving the network at time t because of their dissatisfaction by the ride-sourcing system. The rate of the waiting passengers in Region i with destination in i and j which become assigned at time t are denoted by $W_{ii}^{\text{W-A}}(t)$ and $W_{ij}^{\text{W-A}}(t)$. Parameters $W_{ii}^{\text{W-A}}(t)$ and $W_{ij}^{\text{W-A}}(t)$ are the outputs of the dispatching subsystem.

Finally, the number of the assigned passengers is:

$$\frac{\mathrm{d}p_{ii}^{\mathrm{A}}(t)}{\mathrm{d}t} = W_{ii}^{\text{W-A}}(t) - q_{ii}^{p^{A-}}(t) - b_{ii}(t) \qquad \forall i \in \mathbb{R}, \tag{2.25}$$

$$\frac{\mathrm{d}p_{ij}^{\mathrm{A}}(t)}{\mathrm{d}t} = W_{ij}^{\text{W-A}}(t) - q_{ij}^{p^{A-}}(t) - b_{ij}(t) \qquad \forall i \in \mathbb{R} \text{ and } j \in \mathbb{U}_i, \tag{2.26}$$

where, $q_{ii}^{p^{A-}}(t)$ and $q_{ij}^{p^{A-}}(t)$ are the rate of impatient assigned passengers in Region i with trip destination in regions i and j that cancel their trips. The expiration of reserved time threshold for impatient assigned passengers are set exogenously as stochastic parameters. $q_{ii}^{p^{A-}}(t)$ and $q_{ij}^{p^{A-}}(t)$ are endogenous parameters that reflect one aspect of quality of the dispatching and transfer subsystems. Desired dispatching and transfer subsystems results in lower values of $q_{ii}^{p^{A-}}(t)$ and $q_{ij}^{p^{A-}}(t)$.

2.5 Results

In this section, we investigate the performance of the proposed vehicle-passenger matching method and evaluate the accuracy of the proposed model using a dynamic ride-sourcing benchmark developed in Aimsun microsimulation. In Figure 2.5, the interaction between different modules of the developed benchmark is illustrated. The calibrated Aimsun microsimulation model of the city center of Barcelona (Kouvelas et al. 2017) is plugged into the benchmark. The studied network approximately covers an area of 8.21 squared kilometers containing 1570 sections and 721 junctions. The Aimsun microscopic model includes other travel modes such as normal vehicles and public transport. While they impact the network congestion jointly, their demands are not correlated. The Aimsun microscopic model updates the model state (e.g. position of normal and ride-sourcing vehicles, passengers, buses) every half a second. In the following, firstly, we scrutinize the effects of the proposed vehicle-passenger matching method (see Section 2.3) in comparison with variants of optimal myopic method. Subsequently, we assess the accuracy of the proposed model (see Section 2.4). The results of the designed controller for transferring idle vehicles are presented in the companion article.

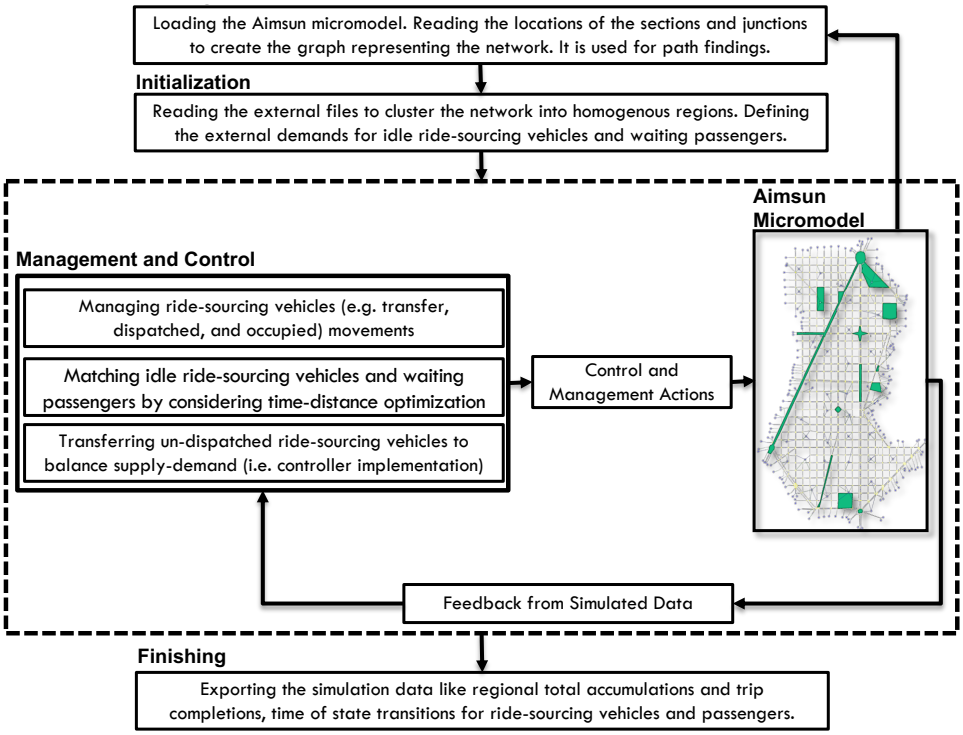


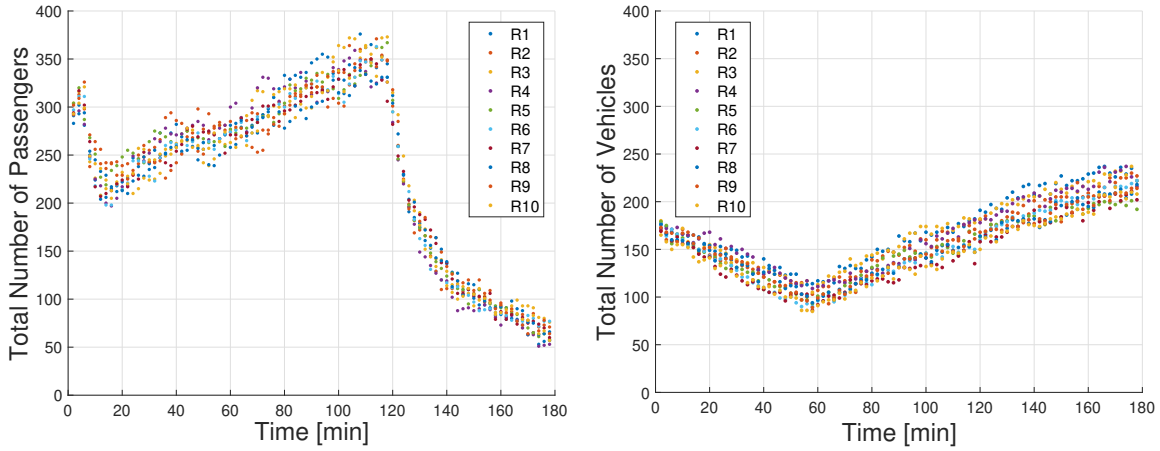
FIGURE 2.5. Schematic interactions between the developed modules of the ride-sourcing microsimulation benchmark.

2.5.1 Analysis of spatio-temporal filtering Method

In this section, the proposed vehicle-passenger matching method is compared with optimal myopic and optimal myopic with greedy discarding under noticeable variation of vehicles supply and passengers demand. In the proposed matching method, spatio-temporal filtering method works in sequence after the optimal myopic matching to prune inefficient (long-distance) vehicle-passenger matchings. Note that the vehicle transferring is not considered in this part to avoid reporting the intertwined effect of transferring with matching methods.

Figure 2.6 illustrates the variation of the total number of passengers and ride-sourcing vehicles in 10 replications by using optimal myopic matching method, without considering impatient vehicles and drivers, with different random initialization for each replication. Furthermore, each replication contains a considerable stochasticity in exogenous demand and supply rates.

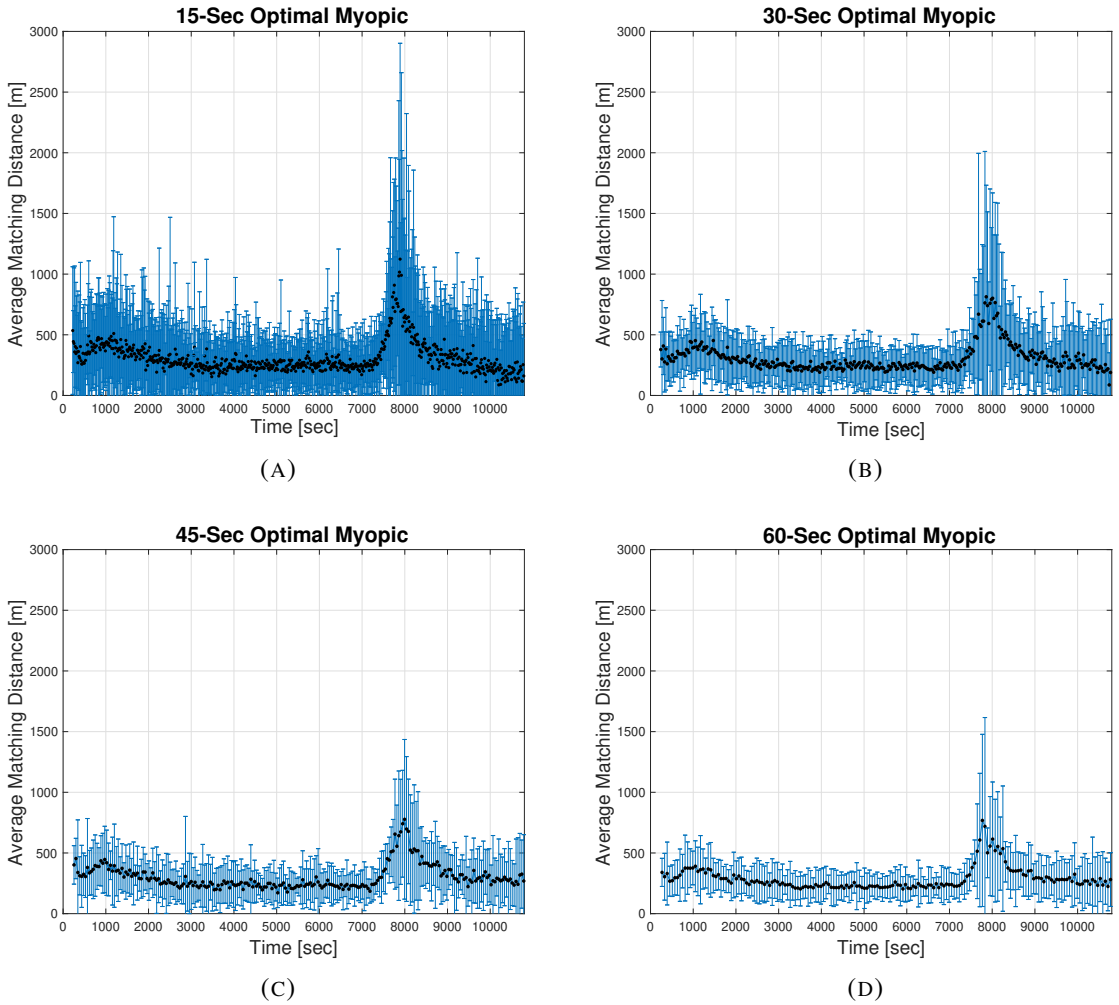
The exogenous demand and supply used in each replication are set to reflect time-varying patterns and under- and over-supply periods.



(A) (B)

FIGURE 2.6. (a) Total number of the passengers including waiting, assigned, and on-board passengers. (b) Total number of the vehicles including idle, dispatched, and occupied vehicles.

Optimal myopic method dispatches vehicles to the passengers based on solving Equation 2.1 in each matching instance without discarding long-distance matchings. This approach considers just the current state of the vehicles and passengers and is sensitive to a predefined matching interval. Moreover, it does not consider the effects of traffic congestion and dynamic of the idle vehicles and waiting passengers. Figure 2.7 depicts the average of matching distances for different matching intervals. Increasing the matching interval results in lower average and variance of vehicle-passenger matching distances (i.e. shorter vehicle's/passenger's reserved time). However, increasing the matching interval increases the unassigned time of vehicles and passengers, see Table 2.1.



(A) (B)

(C) (D)

FIGURE 2.7. Average of matching distances at each matching time using optimal myopic method with different matching interval: (a) 15-second, (b) 30-second, (c) 45-second, and (d) 60-second. The bars represent 95% confidence interval.

Figure 2.8 presents the average of matching distances for matching with greedy discarding method. The greedy method prunes the outcomes of the optimal myopic by discarding k long-distance vehicle-passenger matching if they are above the predefined distance threshold. This approach only takes into account the current number and location of the passengers and vehicles. The illustrated results in Figure 2.8 are obtained by choosing $k = 1$ and distance threshold of 900 [m]. Distance threshold is determined using cut off value of 99% for

considering potential outlier machings, see Figure 2.7. Effectiveness of the greedy method is quantified is Table 2.1.

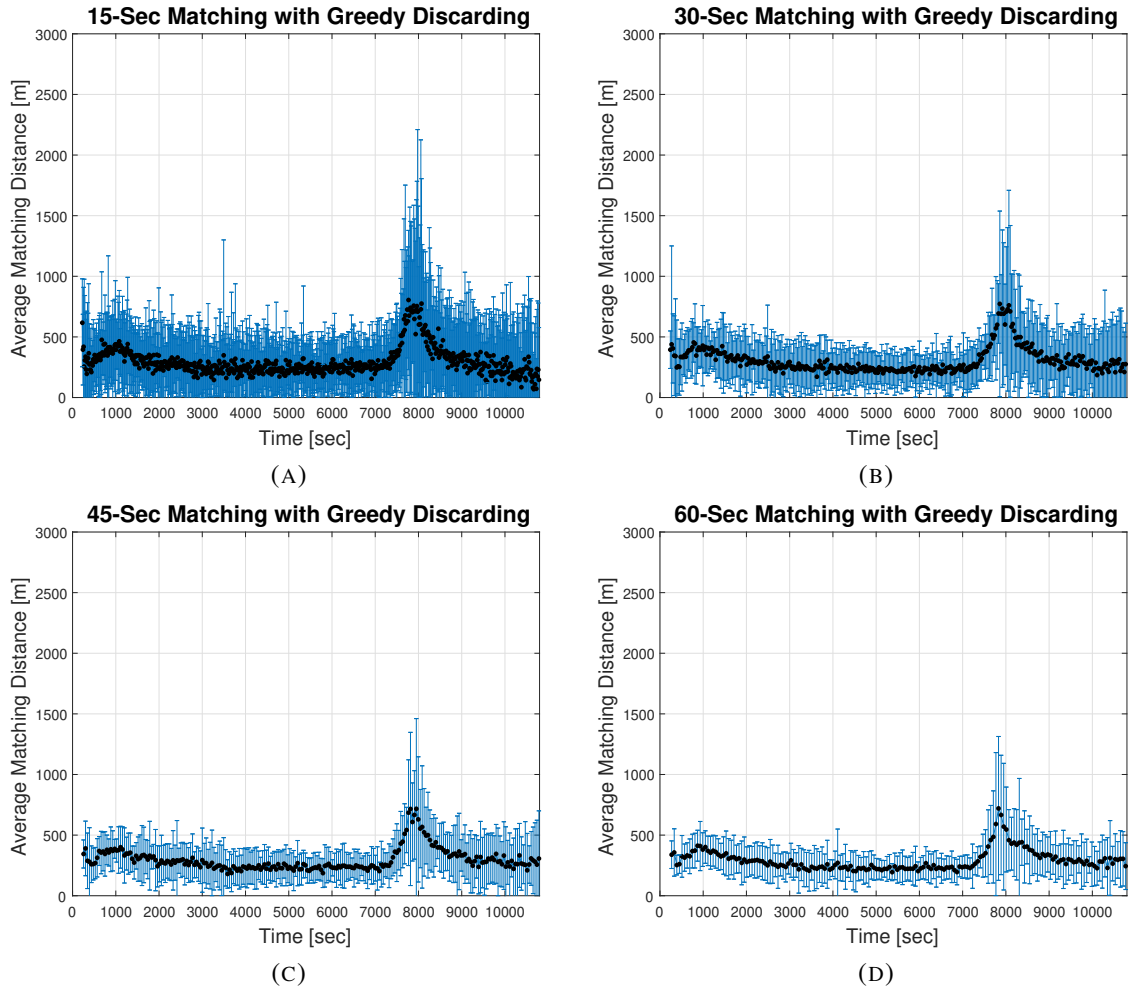
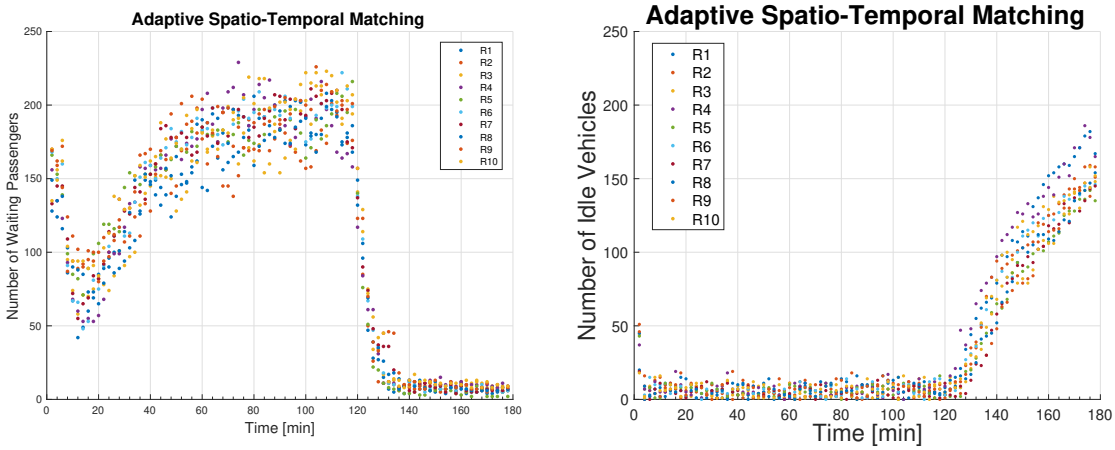


FIGURE 2.8. Average of matching distances in each matching time using greedy discarding method with different matching interval: (a) 15-second, (b) 30-second, (c) 45-second, and (d) 60-second. The bars represent 95% confidence interval.

In the proposed vehicle-passenger matching method, the optimum value of the matching interval and the maximum matching distance are determined by solving Equation 2.7. These optimum values are time-varying such that at each matching instance, the optimum values of the current distance threshold and the time for the next matching instance are obtained. Figure 2.9 illustrates the number of waiting passengers and idle vehicles in the network by

implementing the adaptive spatio-temporal matching method. Figure 2.10 elucidates how matching interval and maximum number of the matching intertwined with each other in the proposed method.



(A) (B)

FIGURE 2.9. (a) Total number of the waiting passengers. (b) Total number of the idle vehicles.

To explore the results of the adaptive spatio-temporal matching method, we discuss about the three-hour simulation results in five time periods: (i) $\Delta t_1 \approx [0 \text{ min}, 12 \text{ min})$, (ii) $\Delta t_2 \approx [12 \text{ min}, 40 \text{ min})$, (iii) $\Delta t_3 \approx [40 \text{ min}, 120 \text{ min})$, (iv) $\Delta t_4 \approx [120 \text{ min}, 140 \text{ min})$, and (v) $\Delta t_5 \approx [140 \text{ min}, 180 \text{ min}]$. In Δt_1, the number of the waiting passengers is decreasing and is greater than the number of the idle vehicles, see Figure 2.9. Hence, the average optimum matching distance based on Equation 2.8 must be increasing as shown in Figure 2.10 (a). The average matching distance has *low-rate* slope because the number of the idle vehicles is *much less* than the number of the waiting passengers. In this time period, the number of the discarded matching is high and increasing, Figure 2.10 (b) and (c), because the density of idle vehicles is low while density of waiting passengers is decreasing. When the number of the discarded matchings is increasing, the matching time interval decreases, Figure 2.10 (d), to compensate the delay caused by discarded matchings in estimated total unassigned time (Equation 2.3) and predicted total reserved time (Equation 2.4).

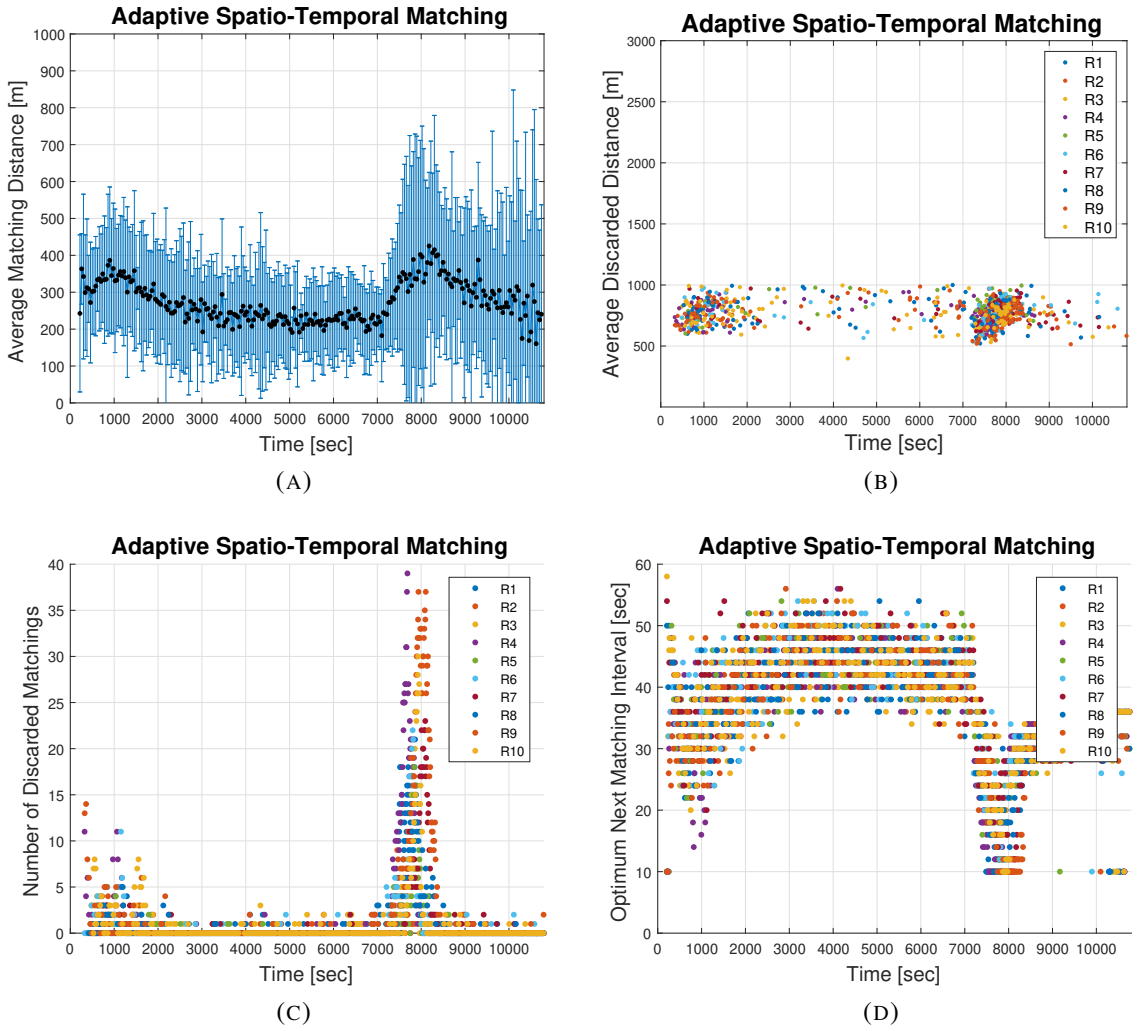


FIGURE 2.10. Output of the adaptive spatio-temporal matching method: (a) average of matching distances after discarding, (b) average distance of discarded matchings, (c) number of discarded matchings, and (d) matching interval. The bars represent 95% confidence interval and different dotted colors show different replications.

In $\Delta t_2 \approx [12 \text{ min}, 40 \text{ min})$, the number of the waiting passengers which is greater than the number of idle vehicles starts to increase while the number of the idle vehicles does not change significantly, see Figure 2.9. Hence, as expected based on Equation 2.8, the average optimum matching distance decreases as in Figure 2.10 (a). The number of the idle vehicles is less than the number of the waiting passengers so the number of the matchings is bounded by the number of the idle vehicles. On the other side, as the number of the

waiting passenger increases, the possibility of short-distance matchings are increased that results in less discarded vehicle-passenger matching as illustrated in Figures 2.10 (b) and (c). Furthermore, the optimum value of the next matching interval is increased, see Figure 2.10 (d), to compensate the effect of increasing the number of the waiting passengers on predicted total unassigned time (see Equation 2.5). Note that the matching method by increasing the matching interval generate fewer discarded matches because waiting passengers are more likely to be matched with idle vehicles in close proximity.

In $\Delta t_3 \approx [40 \text{ min}, 120 \text{ min})$, the number of the idle vehicles and the waiting passengers are not changed notably. Hence, the average of matching distances, number of the discarded matching, and matching intervals are not refined significantly. The trend of the idle vehicles and waiting passengers in $\Delta t_4 \approx [120 \text{ min}, 140 \text{ min})$ are almost the same as Δt_1. Hence, the same explanation is valid. The trend of idle vehicles and waiting passengers in $\Delta t_5 \approx [140 \text{ min}, 180 \text{ min}]$ are similar to Δt_2 if we use use idle vehicle and waiting passengers interchangeably. Because of symmetric characteristic of a matching problem, refer to Equation 2.8 and Figure 2.4 (b), the explanation for Δt_2 can be applied for Δt_5.

Table 2.1 presents the quantitative comparison of different matching methods. In this table, total delay is sum of reserved time and unassigned time for vehicles and passengers as well as the delay of the impatient waiting/assigned passengers and impatient idle vehicle drivers that cancel their trips or leave the ride-sourcing system. Greater matching intervals decrease the average of matching distance that is equivalent to the reduction in the reserved time of vehicles and passengers. However, this inflates the unassigned time of passengers and vehicles. The proposed method shows significant improvement in reducing total delay. It is worth to point out the unassigned time of the vehicles in 15-second optimal myopic and 15-second matching with greedy discarding is less than the proposed method because vehicles unassigned time just considers the vehicles that *successfully* are matched to the waiting passengers. Passengers' unassigned time of the proposed method is in a same level of two other methods because the latter methods assign idle vehicles and waiting passengers together once they become available; however, some of these matchings are long-distance such that some assigned

passengers leave the network before boarding that are not reflected in passengers' unassigned time.

The number of the (impatient) passengers who leave the network before boarding (i.e., order cancellations) for each method is reported in Table 1. These values are obtained by averaging over the 10 replications. The number of the order cancellations by utilizing the adaptive spatio-temporal method is lower than greedy discarding and optimal myopic methods. Note that, in each of the three-hour simulation replication, the total number of trip requests is 2500.

TABLE 2.1. Comparison of adaptive spatio-temporal method and variants of optimal myopic methods with respect to reserved time, unassigned time, total delay, and the number of order cancellations. The matching interval for adaptive spatio-temporal method is time-varying and the reported value is the average of the matching intervals.

	Matching Interval [sec]	Number of Order Cancellation	Passengers' Unassigned Time [sec]		Vehicles' Unassigned Time [sec]		Vehicles/Passengers Reserved Time [sec]		Total Delay [sec]	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
Optimal Myopic (No Discarding)	15	49.9	143.7	134.1	38.6	167.9	93.9	85.5	457.3	183.3
Optimal Myopic (No Discarding)	30	53.3	151.7	137.8	45.5	174.6	91.4	83.1	482.2	196.7
Optimal Myopic (No Discarding)	45	47.0	157.4	139.7	51.6	159.7	90.0	81.4	500.9	184.7
Optimal Myopic (No Discarding)	60	48.4	158.9	139.2	55.7	139.8	87.7	81.3	501.0	172.5
Matching with Greedy Discarding	15	49.5	143.4	135.2	38.2	157.8	91.8	82.7	458.0	183.5
Matching with Greedy Discarding	30	50.9	150.1	136.4	44.8	162.8	91.4	81.7	476.8	182.3
Matching with Greedy Discarding	45	42.4	155.3	138.0	51.2	157.1	88.8	79.8	489.9	181.0
Matching with Greedy Discarding	60	47.1	158.3	138.8	58.3	171.3	86.8	81.7	503.7	183.8
Adaptive Spatio-Temporal Method	34.8 (Mean)	38.8	143.0	128.1	48.1	159.0	80.6	74.3	371.2	158.4

As shown in Figure 2.6, from time 0 [min] to 130 [min], the total number of the ride-sourcing vehicles is less than the total number of the passengers (i.e., under-supply period). The effect of the spatio-temporal matching method during this period is investigated in Table 2.2. In Table 2.3, the effect of the proposed matching when the total number of the ride-sourcing vehicles is greater than the total number of the passengers, i.e. time 130 [min] to 180 [min], is evaluated (over-supply period). The number in parenthesis represent the percentage of changes with respect to average of performances of optimal myopic method variants. The comparison of Table 2.2 and Table 2.3 reveals that the proposed method reduces the number of order cancellations, reserved time, passengers' unassigned time, and total delay in both under- and over-supply periods. Once the market is under-supply, the vehicles' unassigned time is almost equal to the average of performances of optimal myopic methods; however, the passengers' unassigned time is reduced 7.6% and the order cancellation is reduced 20.4%. Noteworthy, the proposed method improves the quality of vehicle-passenger matchings more

during the under-supply period. This supports the conjecture that any matching algorithms is more effective in under-supplied markets; this is because matching algorithms enable efficient use of resources, the supply side.

TABLE 2.2. Comparison of adaptive spatio-temporal method and variants of optimal myopic methods with respect to reserved time, unassigned time, total delay, and number of order cancellation for simulation time from 0 [min] to 130 [min] (i.e., under-supply period). The number in parenthesis represent the percentage of changes with respect to average of performances of optimal myopic method variants.

	Matching Interval [sec]	Number of Order Cancellation	Passengers' Unassigned Time [sec]		Vehicles' Unassigned Time [sec]		Vehicles/Passengers Reserved Time [sec]		Total Delay [sec]	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
Optimal Myopic (No Discarding)	15	29.0	168.4	131.2	18.4	135.1	90.7	82.8	459.0	165.9
Optimal Myopic (No Discarding)	30	36.4	172.0	135.5	27.9	162.2	87.2	80.1	486.7	181.5
Optimal Myopic (No Discarding)	45	31.1	186.1	137.8	32.7	126.1	84.9	77.9	504.6	163.8
Optimal Myopic (No Discarding)	60	29.4	185.6	138.1	39.2	118.7	84.4	75.9	509.4	163.1
Matching with Greedy Discarding	15	34.7	172.6	132.3	18.8	128.7	87.9	79.9	460.7	164.2
Matching with Greedy Discarding	30	35.8	176.7	134.5	26.9	138.3	86.9	77.1	481.0	169.2
Matching with Greedy Discarding	45	28.1	182.2	136.6	33.9	145.0	85.1	75.4	496.4	172.5
Matching with Greedy Discarding	60	29.8	184.0	138.6	40.6	138.2	85.5	74.5	510.5	173.6
Adaptive Spatio-Temporal Method	36.0 (Mean)	25.3 (-20.4%)	164.8 (-7.6%)	130.3	29.8 (0.0%)	145.6	82.9 (-4.2%)	68.6	400.8 (-17.9%)	157.3

TABLE 2.3. Comparison of adaptive spatio-temporal method and variants of optimal myopic methods with respect to reserved time, unassigned time, total delay, and number of order cancellation for simulation time between 130 [min] to 180 [min] (i.e., over-supply period). The number in parenthesis represent the percentage of changes with respect to average of performances of optimal myopic method variants.

	Matching Interval [sec]	Number of Order Cancellation	Passengers' Unassigned Time [sec]		Vehicles' Unassigned Time [sec]		Vehicles/Passengers Reserved Time [sec]		Total Delay [sec]	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
Optimal Myopic (No Discarding)	15	20.9	8.3	6.9	114.7	122.6	111.7	94.3	372.7	144.3
Optimal Myopic (No Discarding)	30	16.9	16.5	14.0	104.0	123.9	112.4	91.8	363.6	146.5
Optimal Myopic (No Discarding)	45	15.9	24.3	16.7	108.1	129.5	114.1	92.0	376.8	148.1
Optimal Myopic (No Discarding)	60	19.0	32.3	22.2	112.2	125.8	103.1	84.0	365.7	142.2
Matching with Greedy Discarding	15	14.8	9.7	11.3	107.1	123.7	111.6	91.0	355.7	142.4
Matching with Greedy Discarding	30	15.1	17.9	15.11	102.7	131.2	113.9	88.2	367.3	151.0
Matching with Greedy Discarding	45	14.3	24.1	16.6	109.6	128.5	106.9	83.0	363.6	140.6
Matching with Greedy Discarding	60	17.3	33.4	24.2	113.9	126.1	104.8	83.7	376.8	150.4
Adaptive Spatio-Temporal Method	33.1 (Mean)	13.5 (-19.5%)	15.9 (-29.3%)	14.7	97.7 (-10.4%)	128.8	93.0 (-4.4%)	72.0	316.8 (-13.8%)	113.2

2.5.2 Model Validation

Traffic congestion is heterogeneous in the case study that results in MFDs with noticeable scatter. As the developed formulations are based on MFDs, we need to partition the network into homogeneous regions to achieve well-defined low-scatter MFDs. Figure 2.11 (a) depicts the partitioning of the network into four homogeneous regions as suggested in (Kouvelas et al. 2017). To validate the proposed model, we consider inter-regional flows for ride-sourcing

vehicles (i.e. transfer, dispatched, and occupied) and passengers (i.e. waiting and assigned) between neighbor regions because the matching algorithm prevents long-distance (i.e. far away regions) matchings between idle/transfer ride-sourcing vehicles and waiting passengers. In addition, the travel requests are defined for the neighbor regions. This assumption is not very limiting for the case study microsimulation network as Regions 1, 2, and 3 are not directly connected. Region 4 is the central region and the other regions are peripheral. We consider the ride-sourcing travel requests are from the central region to the peripheral regions and vice versa. However in real practice with different topology of regions, a general form of origins and destinations of ride-sourcing trips should be considered. This would require defining more traffic states, which is a straightforward yet cumbersome extension of the proposed macro model.

We excite the microsimulation model with different demand levels and initial states to obtain the regional MFDs. Figures 2.11 (b) to 2.11 (e) illustrates the simulated and estimated MFD functions. Different colors distinguish the results of different replications. The solid lines represent the estimated MFDs with constrained least square method based on the simulated data. We utilize the estimated MFDs in evaluating the proposed ride-sourcing model.

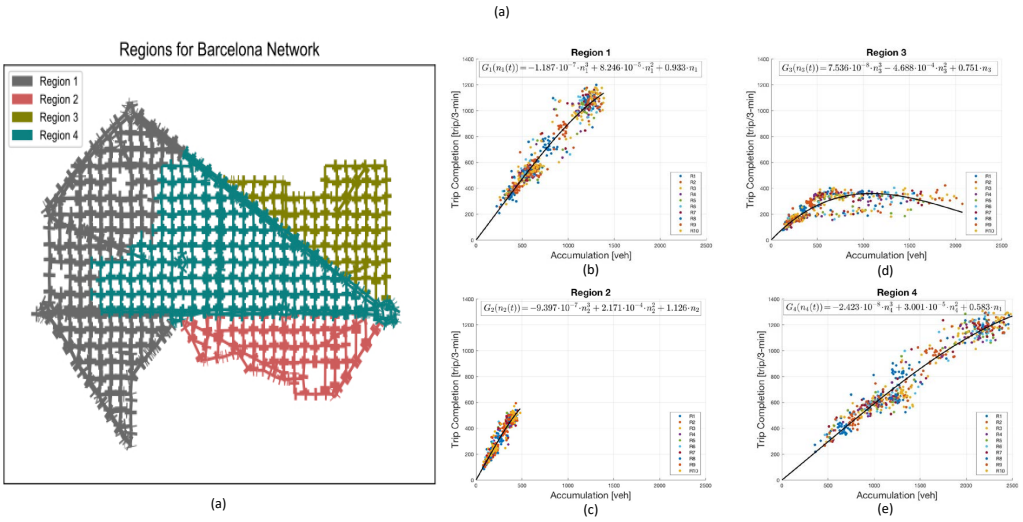


FIGURE 2.11. (a) Homogeneous regions of the city center of Barcelona. Macroscopic Fundamental Diagram (MFD) for (b) Region 1, (c) Region 2, (d) Region 3, and (e) Region 4. n_i denotes the total accumulation including buses, ride-sourcing vehicles, and normal vehicles in Region i. Trip completion rate in Region i is denoted by G_i. The solid lines show the estimated MFD functions. The dotted colors illustrates the simulation results of different replications.

Figures 2.12 to 2.16 assess the accuracy of the proposed macroscopic formulation using the adaptive spatio-temporal matching method as the dispatching subsystem. This model validation aims to study the accuracy of macroscopic aggregation of boarding rate with Cobb-Douglas meeting function in Equation 2.9, homogeneity assumption in Equation 10, and MFD dynamics in Equations 2.11 to 2.15. Figure 2.12 depicts the estimated number of the waiting passengers (Equations 2.23 and 2.24) and simulated number of the waiting passengers with dotted red lines. Root mean square error (RMSE) of estimated waiting passengers for regions 1 to 4 are: 1.51, 0.88, 0.82, and 2.17, respectively. The estimated values closely follow the simulated values because the exogenous terms are dominant in estimating the number of waiting passengers in each region.

For model validation that is reported in these figures, two sets of information are fed to the model in addition to the exogenous inputs (i.e., the arrival rate of idle vehicles and waiting passengers): initial values of states, and the rate of vehicle-passenger matchings. The main sources of uncertainties of the model are (i) initial location of new arriving idle vehicles and new passengers in each region, (ii) stochastic thresholds of patience time of passengers and drivers, and (iii) cruising movements of idle vehicles. Figures 11 and 15 shows higher accuracy as the evolution of the waiting passengers and idle vehicles are dominated by exogenous inputs. In Figures 13 and 14, the abrupt variation of the occupied vehicles cannot be capture as the evolution of these sates are estimated dominantly by MFD functions that are more adapt for slow-varying systems.

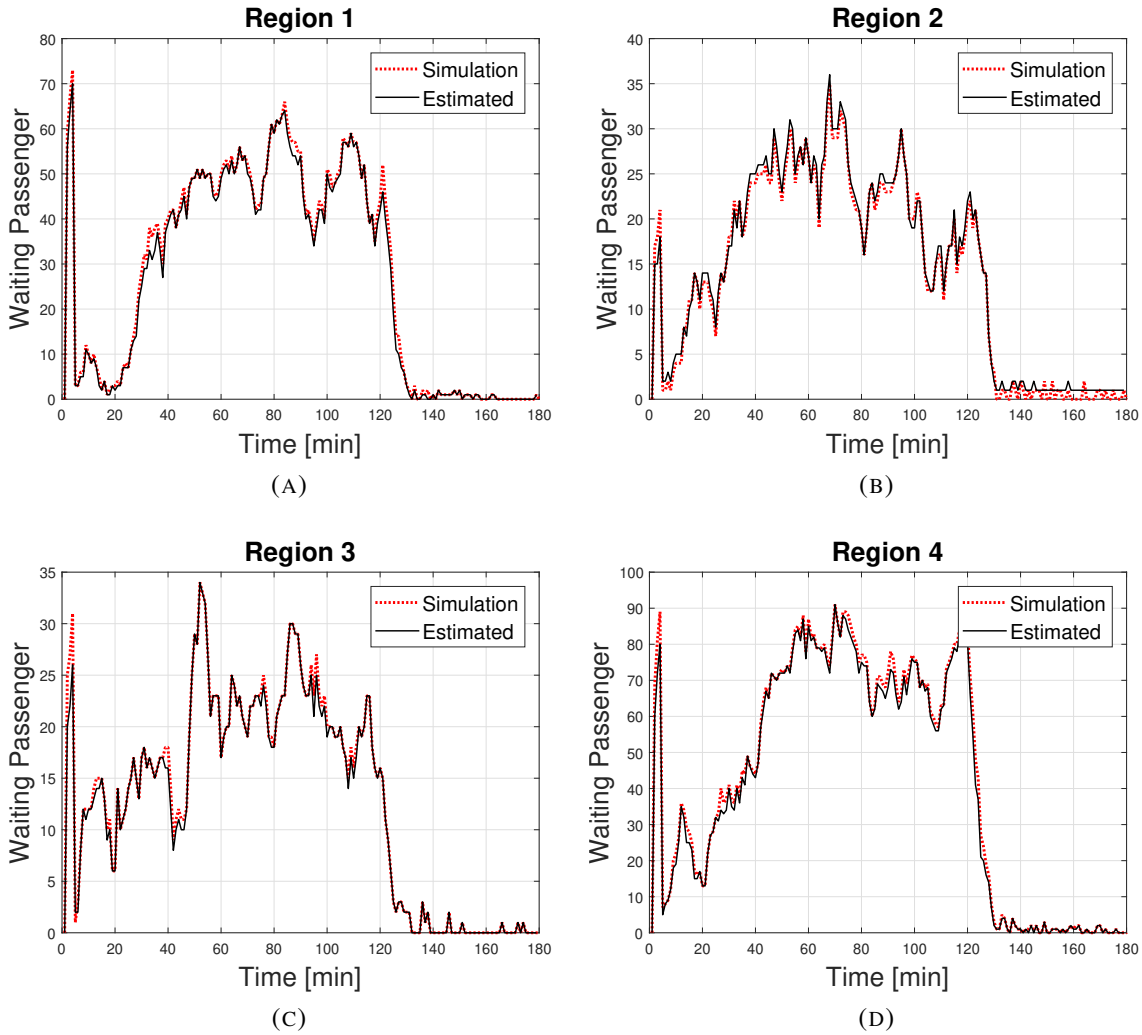


FIGURE 2.12. Comparison of estimated and simulated evolution of the number of waiting passengers for: (a) Region 1, (b) Region 2, (c) Region 3, and (d) Region 4. The black solid lines show the estimated values of regional waiting passengers. The red dotted lines illustrate the simulation results.

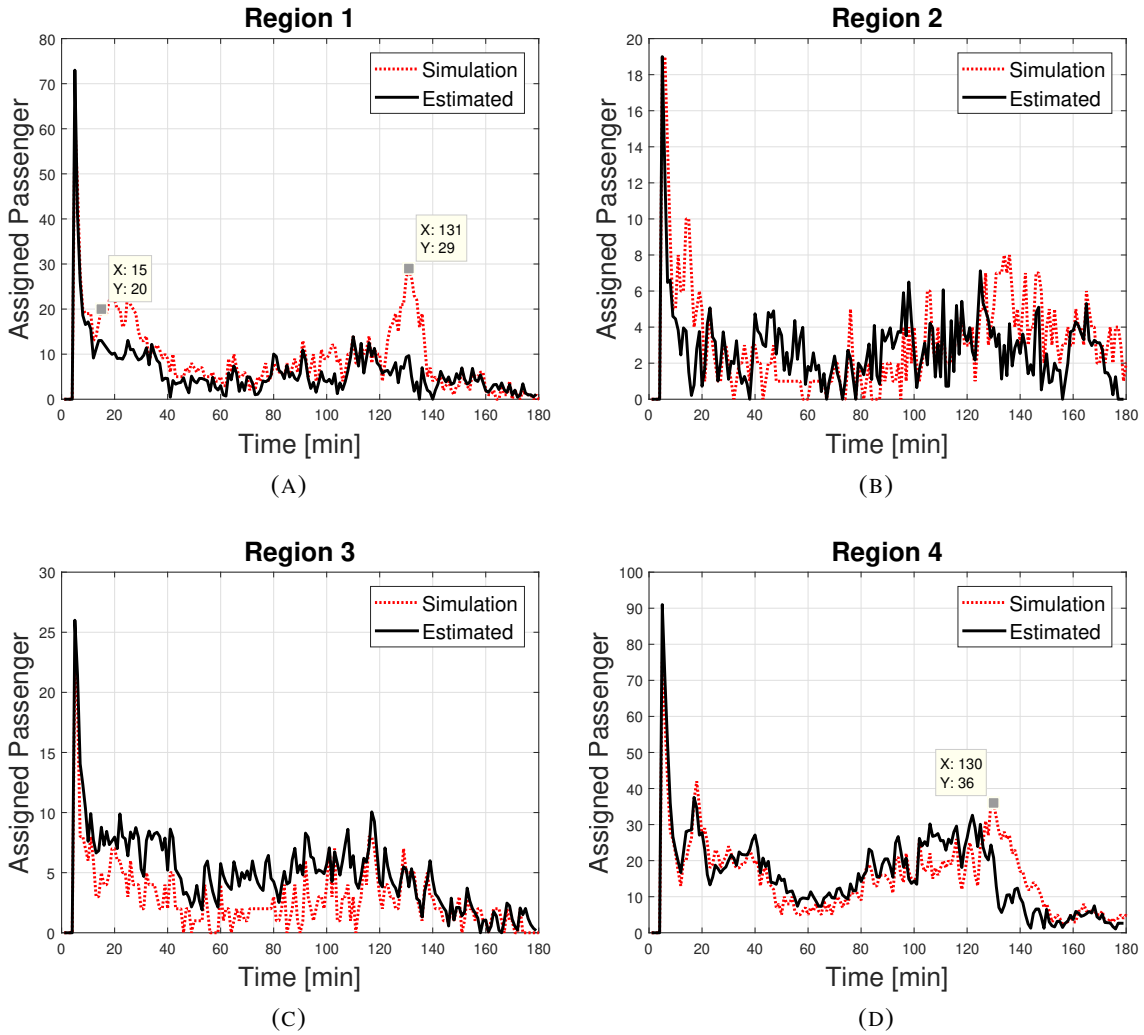


FIGURE 2.13. Comparison of estimated and simulated evolution of the number of assigned passengers for: (a) Region 1, (b) Region 2, (c) Region 3, and (d) Region 4. The black solid lines show the estimated values of regional assigned passengers. The red dotted lines illustrate the simulation results.

Estimated number of assigned passengers, Equations 2.25 and 2.26, and simulated number of assigned passengers in regions one to four are compared in Figure 2.13. The initial peaks in this Figure refer to high number of idle vehicles and waiting passengers that are considered in matching procedure at the beginning of the simulation. The RMSE of estimated number of assigned passengers of regions 1 to 4 are: 5.66, 2.61, 2.59, and 8.42. Cobb-Douglas matching function that is used in Equations 2.25 and 2.26 cannot capture abrupt changes near extrema

(e.g. time 15 [min] and 130 [min] where there are minimum and maximum extrema in total number of the passengers as shown in Figure 2.6.

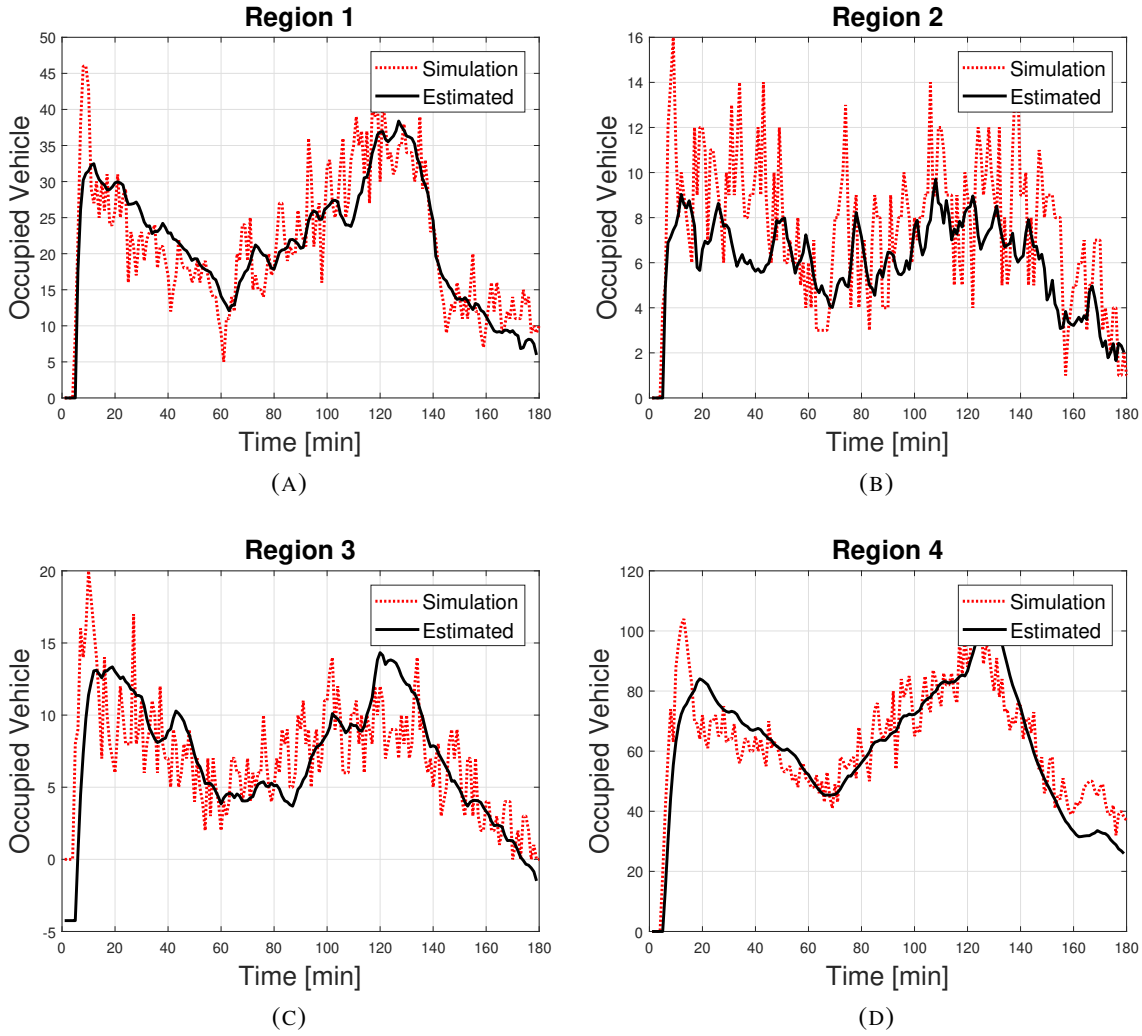


FIGURE 2.14. Comparison of estimated and simulated evolution of the number of occupied vehicles for: (a) Region 1, (b) Region 2, (c) Region 3, and (d) Region 4. The black solid lines show the estimated values of regional occupied vehicles. The red dotted lines illustrate the simulation results.

RMSE for estimated number of occupied vehicles, Equations 2.16 and 2.17, in regions 1 to 4 are: 4.86, 3.04, 3.24, and 8.57. The occupied vehicles intrinsically are slow-varying mode of ride-sourcing systems so they can be estimated using MFDs and Cobb-Douglas functions accurately. RMSE for estimated dispatched vehicles, Equations 2.18 and 2.19, in regions 1

to 4 are: 4.20, 8.21, 1.57, and 6.31. The estimated and the simulated number of dispatched vehicles are depicted in Figure 2.15. Similar to assigned passengers, the dispatched vehicles are dominated by the results of dispatching subsystems that can impose high variation trends.

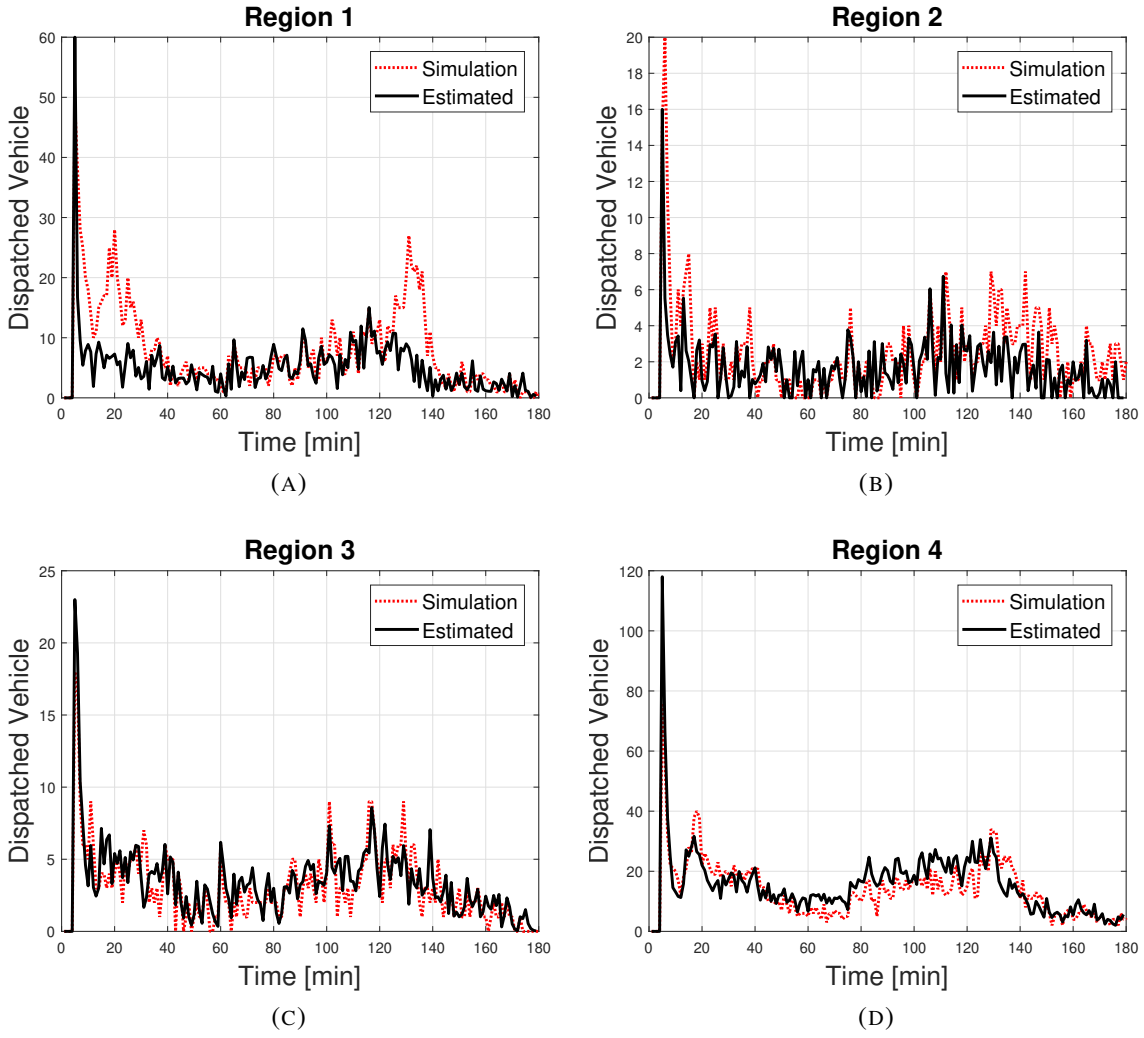


FIGURE 2.15. Comparison of estimated and simulated evolution of the number of dispatched vehicles for: (a) Region 1, (b) Region 2, (c) Region 3, and (d) Region 4. The black solid lines show the estimated values of regional dispatched vehicles. The red dotted lines illustrate the simulation results.

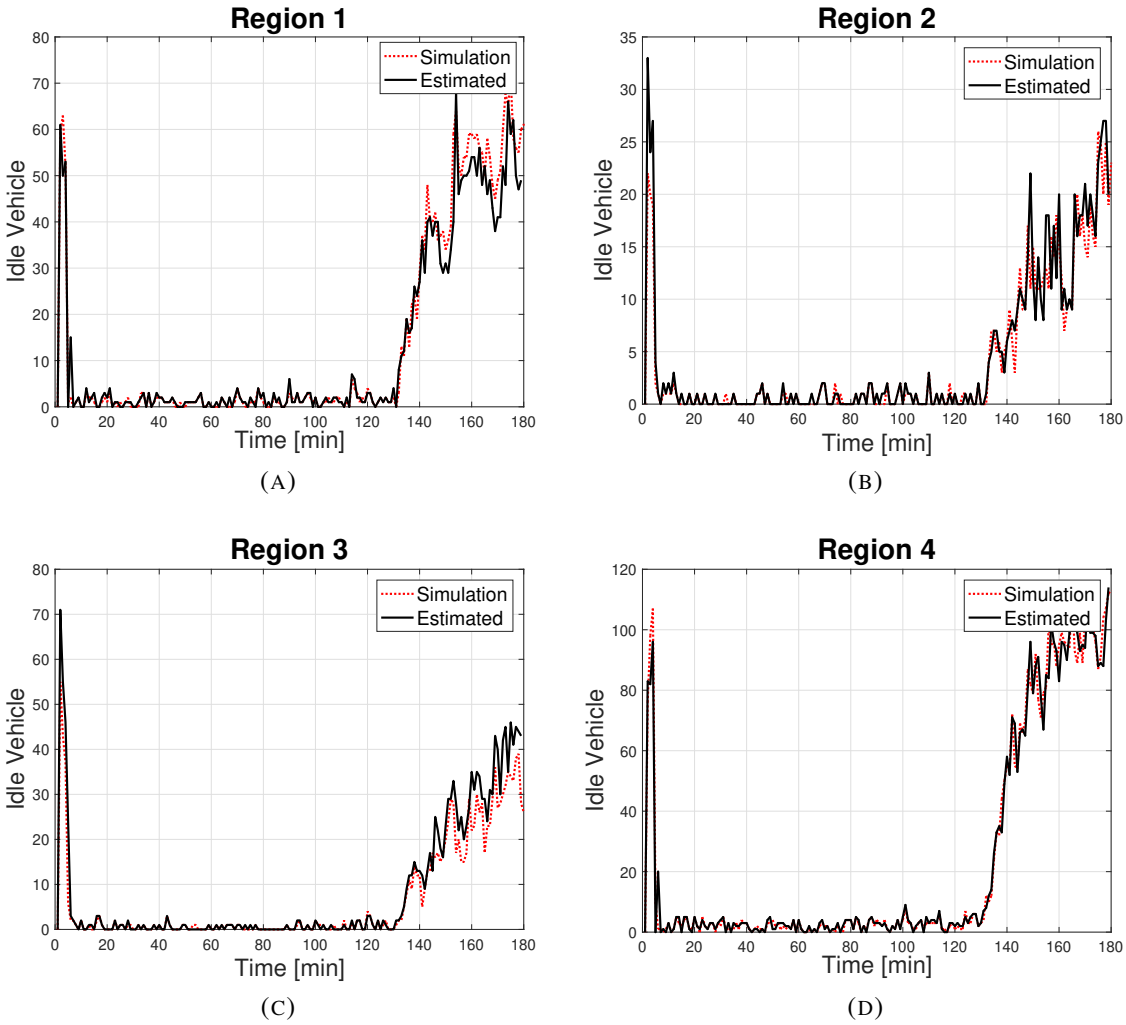


FIGURE 2.16. Comparison of estimated and simulated evolution of the number of idle vehicles for: (a) Region 1, (b) Region 2, (c) Region 3, (d) Region 4. The black solid lines show the estimated values of regional idle vehicles. The red dotted lines illustrate the simulation results.

Figure 2.16 illustrates the comparison of the estimated idle vehicles, Equation 2.22, and simulated idle vehicles. RMSE for estimation of idle vehicles are: 3.93, 2.08, 4.06, and 4.39

2.6 Summary and Future Research

This article has presented a dynamic macroscopic model and a matching method for ride-sourcing systems. The macroscopic dynamical ride-sourcing model is developed based on partitioning the network into a set of homogeneous regions and tracks the spatio-temporal evolution of different states of ride-sourcing vehicles (i.e. transferred, idle, dispatched, and occupied) and passengers (i.e. waiting and assigned) between regions. The model built upon the Macroscopic Fundamental Diagrams (MFDs) and Cobb-Douglas matching function to derive the non-equilibrium macroscopic model. Contrary to the conventional approaches, we proposed a non-equilibrium model that overcomes the limitation of stationary and steady-state analysis of the ride-sourcing systems. The model provides the capabilities of considering vehicle-passenger matching scenarios and designing a controller for repositioning of idle ride-sourcing vehicles. The proposed macroscopic model is evaluated with micro-level data utilizing a microsimulation benchmark of ride-sourcing systems.

A Nonlinear Model Predictive Controller is developed in Chapter 3 based on the framework and the model that is suggested in this study. Impatient passengers and drivers are considered in the proposed model endogenously. In addition to the ride-sourcing modeling, we proposed an algorithm to dynamically determine the optimum matching intervals and maximum matching distance with respect to minimizing passengers' waiting time. The algorithm considered (i) the intertwine effect of matching time interval and maximum matching distance, (ii) level of congestion of the network, and (iii) dynamics of waiting passengers and idle/transferred vehicles to find the optimum values at each matching time instance. The validity of the macroscopic model and the benefits of the matching method have been demonstrated with microsimulation.

Several future research directions are envisaged. The passengers and ride-sourcing vehicles arrival profiles can be modeled elastic (in short-term) to the market conditions (e.g. passengers' and vehicle's waiting time). This will lead to positive and negative short-term induced demand and supply that requires more behavioural research on both sides of the market. Furthermore, speed as an indicator of network congestion is considered as an exogenous factor. Integrating

estimation of the network congestion indicator from the ride-sourcing fleet (e.g. as probe vehicles) is another future research direction. Moreover, effects of competition between different ride-sourcing companies on vehicle-passenger matching and repositioning is a challenging future study. The proposed matching method relies on solving the matching problem of a bipartite graph. Scaling of this to obtain the exact solution is a challenge that requires further research (Bertsimas et al. 2019). Furthermore, extending the proposed framework, method, and model to ride-splitting (ride-sharing), is a research priority.

Appendix: Nomenclature

Sets

$\mathbb{R}$	set of homogeneous regions
$\mathbb{U}_i$	set of regions in the vicinity of Region i

Functions

$b_{ij}(t)$	vehicle boarding rate from Region i to Region j at time t
P_i	total trip production (MFD) in Region i

Parameters

$l_{ij}^{\mathrm{O}}(t)$	average trip length of occupied vehicles in i with destination in j at time t
$l_{ij}^{\mathrm{T}}(t)$	average trip length of transferred vehicles in i with destination in j at time t
$l_{ij}^{\mathrm{D}}(t)$	average trip length of dispatched vehicles in i with destination in j at time t
α_i	boarding function elasticity with respect to dispatched vehicles in i
β_i	boarding function elasticity with respect to assigned passengers in i
γ_i	boarding function elasticity with respect to average speed of i
K_i	total productivity of boarding function in i

Variables

$\hat{T}_{\mathrm{R}}(t_{\mathrm{m}}^{\iota})$	estimated total reserved time of matchings at t_{m}^{ι}
$\hat{T}_{\mathrm{D}}(t_{\mathrm{m}}^{\iota})$	estimated total unassigned time of the waiting passengers at t_{m}^{ι}

$\hat{T}_{\mathrm{R}}(t_{\mathrm{m}}^{\iota+1})$	predicted total reserved time for matchings at $t_{\mathrm{m}}^{\iota+1}$
$\hat{T}_{\mathrm{D}}(t_{\mathrm{m}}^{\iota+1})$	predicted total unassigned time for waiting passengers at $t_{\mathrm{m}}^{\iota+1}$
$\hat{T}_{\mathrm{P}}^{t_{\mathrm{m}}^{\iota+1}-t_{\mathrm{m}}^{\iota}}$	expected total passengers' waiting time during interval $[t_{\mathrm{m}}^{\iota}, t_{\mathrm{m}}^{\iota+1})$
$m(t_{\mathrm{m}}^{\iota})$	number of matchings before discarding of long-distance matchings at t_{m}^{ι}
$\bar{l}_r(t_{\mathrm{m}}^{\iota})$	average distance of optimal matchings after discarding wild goose chase at t_{m}^{ι}
$\hat{\bar{l}}_r(t_{\mathrm{m}}^{\iota+1})$	prediction of average matching distance
$r(t_{\mathrm{m}}^{\iota})$	number of long-distance vehicle-passenger discardings at t_{m}^{ι}
$\rho_{\mathrm{c}}(t_{\mathrm{m}}^{\iota})$	total arrival rate of idle/transferred vehicles during interval $[t_{\mathrm{m}}^{\iota}, t_{\mathrm{m}}^{\iota+1})$
$\rho_{\mathrm{p}}(t_{\mathrm{m}}^{\iota})$	total arrival rate of waiting passengers during interval $[t_{\mathrm{m}}^{\iota}, t_{\mathrm{m}}^{\iota+1})$
$t_{\mathrm{m}}^{\max}$	upper bound for the next matching time instance
$c_{ij}^{\mathrm{O}}(t)$	number of occupied vehicles in Region i with destination in Region j at time t
$c_{ij}^{\mathrm{T}}(t)$	number of transferred vehicles in Region i with destination in Region j at time t
$c_{ij}^{\mathrm{D}}(t)$	number of dispatched vehicles in Region i with destination in Region j at time t
$c_{i}^{\mathrm{I}}(t)$	number of idle vehicles in Region i time t
$p_{ij}^{\mathrm{W}}(t)$	number of waiting passengers in Region i with destination in Region j at time t
$p_{ij}^{\mathrm{A}}(t)$	number of assigned passengers in Region i with destination in Region j at time t
$M_{ij}^{\mathrm{O}}(t)$	occupied vehicle inter-region flow between Region i and Region j at time t
$M_{ij}^{\mathrm{T}}(t)$	transferred vehicle inter-region flow between Region i and Region j at time t
$M_{ij}^{\mathrm{D}}(t)$	dispatched vehicle inter-region flow between Region i and Region j at time t
$M_{ii}^{\mathrm{O}}(t)$	occupied vehicle internal flow in Region i at time t
$M_{ii}^{\mathrm{T}}(t)$	transferred vehicle internal flow in Region i at time t
$v_i(t)$	average speed of Region i
$\hat{v}(t_{\mathrm{m}}^{\iota})$	estimated network average speed at time t_{m}^{ι}
$W_{ij}^{\text{I-D}}(t)$	rate of idle vehicles in Region i that dispatches to Region j
$W_{ij}^{\text{T-D}}(t)$	rate of transferred vehicles in Region i that dispatches to Region j
$W_{ij}^{\text{I-T}}(t)$	rate of idle vehicles in Region i that transfers to Region j
$W_{ij}^{\text{W-A}}(t)$	rate of the waiting passengers which become assigned in i to j at time t
R_{ij}^{D}	rate of the cancellation of dispatched trips in i with destination in j at time t
$w_i^{\mathrm{I}}(t)$	rate of the idle vehicles that enter/leave i at time t
$q_i^{c^+}(t)$	rate of idle vehicles exogenously enters to the network in i at time t

$q_i^{c^-}(t)$	rate of idle ride-sourcing vehicles leave the network from i at time t
$q_{ij}^{p^{W+}}(t)$	exogenous rate of waiting passenger demand in i going to Region j at time t
$q_{ij}^{p^{W-}}(t)$	leaving rate of impatient unassigned passengers in i with destination in j at time t
$q_{ij}^{p^{A-}}(t)$	leaving rate of impatient assigned passengers in i with destination in j at time t

CHAPTER 3

Proactive Vehicle Repositioning

Emergence of ride-sourcing systems with internet-based platforms introduces a popular transportation mode in recent years. The service quality of the ride-sourcing systems and their effects on transportation networks rely on effective matching and *redistributing of idle vehicles*. In Chapter 2, we introduced a vehicle-passenger matching that consider the joint effect of the maximum matching distance and matching intervals to minimize passengers' waiting time. Moreover, a macroscopic non-equilibrium based model is developed to predict the evolution of the states of the ride-sourcing system. In this Chapter, we propose a control method to transfer idle vehicles to balance the demand of waiting passengers and supply of ride-sourcing vehicles by positioning the idle vehicles in locations where there is a higher possibility of faster pickups of new passengers. We develop a proactive controller in which a Nonlinear Model Predictive Controller (NMPC) predicts the future state of the ride-sourcing system determined by the developed non-equilibrium model of Chapter 2. The proposed NMPC implements an optimum rolling horizon strategy as a result of solving a constrained optimization problem at each transferring sample time. The effect of the network congestion and impatient passengers and drivers that are considered in the model is reflected in the proposed method. We show the effectiveness of the designed controller with respect to reducing passengers' unassigned and waiting times, reducing vehicles' waiting times, reducing the number of the fleet size, and increasing the number of the served trip requests by microsimulation experiments. We show that the pro-activeness of the NMPC incentivises the vehicles' drivers to follow the transferring commands. In microsimulation experiments, the designed controller improves the performance of the ride-sourcing system by reducing passengers' average unassigned time (-20.4%) and waiting times (-12.4%), vehicles' average waiting times (-8.8%), the number of the fleet size (-18.6%), and increasing the number of the served trip requests (9.7%).

3.1 Introduction

The emergence of ride-sourcing systems unprecedentedly have changed the market of on-demand mobility services. There has been a continuing debate/research about the effect of dispatching methods of ride-sourcing systems on network congestion, deadheading time of the vehicles, and waiting time of the passengers. The dispatching method is a critical component of the ride-sourcing systems that assign idle vehicles to waiting passengers (i.e. travel requests). However, dispatching methods have the limitation of not manipulating the location of unmatched vacant vehicles. In this study, we propose a dynamic transferring method for repositioning of idle vehicles to the locations with higher probability of being assigned to new emerging passengers. A Nonlinear Model Predictive Controller (NMPC) is designed as a proactive transferring method to reduce the (i) passengers' unassigned and waiting times, (ii) vehicles' waiting times, and (iii) guided vacant cruising distances. We utilize the developed macroscopic non-equilibrium model introduced in Chapter 2 in designing the controller.

Ride-sourcing systems have been studied from different perspectives such as pricing of the service (Zuniga-Garcia et al. 2020; Zha et al. 2018b; Nourinejad and Ramezani 2019; Battifarano and Qian 2019; Zha et al. 2018a), demand or supply estimation (Wu et al. 2018; Ke et al. 2017; Ke et al. 2018), fleet size management (Vazifeh et al. 2018), and vehicle-passenger matching method (Yang et al. 2020; Maciejewski et al. 2016). Interested readers are referred to (Narayanan et al. 2020; Wang and Yang 2019) for detailed reviews on ride-sourcing systems .

A group of articles in literature study the repositioning of idle vehicles without considering dynamic evolution of the system (i.e., reactive repositioning) or between specific stations/nodes. For example (Pavone et al. 2012; Zhang and Pavone 2016; Braverman et al. 2019) propose repositioning methods by assuming stationary conditions. (Sayarshad and Chow 2017) studies a non-myopic policy for idle vehicle repositioning between a set of stations/nodes using queue delay as an approximation of the conditional expected cost. In (Alonso-Mora et al. 2017), a scalable method is suggested for relocating the idle vehicles to the regions with

high instantaneous demand. A balancing method is studied in (Ramezani and Nourinejad 2018b) for a street-hailing taxi system based on macroscopic fundamental diagrams (MFDs). Demand forecasting is incorporated into the paradigm of repositioning of the on-demand mobility systems in (Dandl et al. 2019). One-way vehicles relocation among the fixed nodes is studied in (Illgen and Höck 2019; Nourinejad et al. 2015).

Reactive repositioning of idle vehicles does not consider the intertwined effects among transferring the idle vehicles at the current time, future state of vehicle-passenger matchings, and vehicle and waiting passengers arrivals. Reactive transferring mainly suffers from increasing the deadheading time of the idle vehicles, increasing the dead-mileage, and wild goose chases in expense of the passengers' waiting time. Moreover, relocating the idle vehicles between specific nodes/stations is applicable for shared mobility systems such as on-way car-sharing and car-pooling with limitations in ride-sourcing systems due to spatial-varying pattern of passengers' arrivals in the network. To fill the gap in literature, we develop a proactive repositioning method based on NMPC for ride-sourcing systems that is beneficial for vehicles' drivers to follow the repositioning recommendations of the systems to reduce the waiting time of the passengers.

In Chapter 2, an adaptive spatio-temporal matching method is proposed to dynamically determine the optimum values of maximum matching distance (to discard long distance vehicle-passenger matchings) and the next matching instance (to establish a time-varying matching frequency). The developed method considers the joint effects of matching interval and maximum matching distance to minimize passengers' waiting time and to avoid wild goose chase. In this Part, we focus on repositioning of the idle vehicles to curb the possible mismatch between supply and demand in the system. To this end, we propose a NMPC as a systematic proactive repositioning method for balancing the idle vehicles and waiting passengers in a network that is partitioned into a number of regions. This study develop a real-time scalable repositioning method that considers the prediction of the dynamic evolution of the ride-sourcing system's state (e.g. idle/ dispatched/ occupied/ transferred vehicles, waiting/ assigned/ on-board passengers) and vehicle-passenger matching rates in each region based on the developed model in Chapter 2. The proposed method considers variation of

demand (i.e. passengers' trip requests) and supply (i.e. ride-sourcing vehicles) as well as the effect of estimated network congestion, impatient drivers, and impatient passengers. We show that the pro-activeness of the NMPC alleviates the wild goose chase problem by reducing dispatching distances of the vehicles that are transferred to the recommended locations. The effectiveness of the proposed method is scrutinized with respect to reducing passengers' and vehicles' total delay, increasing the number of the served passengers' trip requests, and reducing operational fleet size.

This article is organized as follows. Section 3.2 elaborates the framework of the dynamic vehicle repositioning system and introduces the mathematical formulation of the transferring system in a state-space form. In Section 3.3, we propose a nonlinear model predictive controller (NMPC) that is built upon solving a rolling-horizon constrained optimization problem. Further, We discuss the estimation of the endogenous inputs for improving the ride-sourcing system's controllability. In Section 3.4, we present the microsimulation setup, a reactive controller, and the results of the transferring method on a traffic microsimulation model. The Chapter is summarised and the future works are discussed in Section 3.5.

3.2 Vehicle Repositioning Framework and Formulation

3.2.1 Framework

The developed ride-sourcing system has two main components: dispatching (introduced in Chapter 2 Chapter) and transfer subsystems. We consider ride-sourcing vehicles cannot street-hail and they should be dispatched to the waiting passengers for pickup. The dispatching subsystem assigns waiting passengers to the vacant vehicles (i.e. idle and transferred vehicles) in each matching instance. The transfer subsystem is based on the partitioning of the network into a set of regions and using the developed model in Chapter 2 to predict the evolution of the idle, transferred, dispatched, and occupied vehicles as well as waiting and assigned passengers. It sends the idle vehicles to the regions with an excess number of waiting passengers to balance the demand of passengers' travel requests with idle vehicles in each region of the network. We

consider idle vehicles cruise around (randomly) before being matched or transferred by the ride-sourcing system. In each *transfer time instance* (i.e. time instances when the transferring subsystem is triggered), the transferring subsystem selects a part of idle vehicles and transfers them to the regions with a high possibility of finding passengers in the next matching instances. Selected idle vehicles follow the routing recommendation of the ride-sourcing system.

The transfer subsystem determines the inter-regional movements of a set of the idle ride-sourcing vehicles based on real-time information of the network and the developed mathematical model. To design the transfer subsystem, in each transferring time instance, we obtain the number of (i) idle vehicles in every region, (ii) transfer, dispatch, and occupied vehicles in Region i with the destination in Region j, (iii) waiting and assigned passengers in Region i with destination in Region j. In addition, the transfer subsystem at each transfer time instance requires an estimation of the rate of (i) incoming idle vehicles in each region and (ii) new waiting passengers in Region i with destination in Region j.

The transfer subsystem utilizes the above information periodically to predict the macroscopic behavior of the network (i.e. evolution of different state of the ride-sourcing system). The predictions are used to determine the rate to transfer the idle vehicles to the regions with a higher possibility of matching to the waiting passengers.

3.2.2 Model Formulation

In this part, first, the summary of the model that is developed in Chapter 2 is briefly re-introduced. In the second part of this section, the state-space representation of the model that is utilised to design the repositioning controller is formulated.

3.2.2.1 Summary of the Model

In this subsection, we briefly summarize the developed macroscopic ride-sourcing model in Chapter 2. The model utilizes the Cobb-Douglas meeting function and Macroscopic Fundamental Diagram (MFD) to represent the evolution of idle, transferred, dispatched, and occupied vehicles as well as waiting and assigned passengers in each region of the network.

In Equations 3.1 to 3.11, subscript ij denotes from current Region i to final destination j. Set of the regions of the network is denoted by $\mathbb{R}$ and $\mathbb{U}_i$ is the set of regions that are in the direct vicinity of Region i.

Evolution of the occupied vehicles, c_{ii}^{O} and c_{ij}^{O}, are formulated in Equations 3.1 and 3.2. The number of dispatched vehicles that become occupied at time t is denoted by $b_{ii}(t)$ and $b_{ij}(t)$.

$$\frac{\mathrm{d}c_{ii}^{\mathrm{O}}(t)}{\mathrm{d}t} = b_{ii}(t) + \sum_{j \in \mathbb{U}_i} M_{ji}^{\mathrm{O}}(t) - M_{ii}^{\mathrm{O}}(t) \qquad \forall i \in \mathbb{R}, \tag{3.1}$$

$$\frac{\mathrm{d}c_{ij}^{\mathrm{O}}(t)}{\mathrm{d}t} = b_{ij}(t) - M_{ij}^{\mathrm{O}}(t) \qquad \forall i \in \mathbb{R} \text{ and } j \in \mathbb{U}_i. \tag{3.2}$$

Equations 3.3 and 3.4 show the conservation equations of dispatched vehicles, $c_{ii}^{\mathrm{D}}(t)$ and $c_{ij}^{\mathrm{D}}(t)$, where $W_{ii}^{\text{I-D}}(t)$ and $W_{ij}^{\text{I-D}}(t)$ are the rate that the idle vehicles become dispatched at time t. The rate of the transferred vehicles that become dispatched are denoted by $W_{ii}^{\text{T-D}}(t)$ and $W_{ij}^{\text{T-D}}(t)$. Inter-regional flow of dispatched vehicles is denoted by $M_{ij}^{\mathrm{D}}(t)$. Furthermore, $R_{ii}^{\mathrm{D}}(t)$ and $R_{ij}^{\mathrm{D}}(t)$ are cancellation rates of the dispatched vehicles by impatient passengers at time t.

$$\frac{\mathrm{d}c_{ii}^{\mathrm{D}}(t)}{\mathrm{d}t} = W_{ii}^{\text{I-D}}(t) + W_{ii}^{\text{T-D}}(t) + \sum_{j \in \mathbb{U}_i} M_{ji}^{\mathrm{D}}(t) - b_i(t) - R_{ii}^{\mathrm{D}}(t) \qquad \forall i \in \mathbb{R}, \tag{3.3}$$

$$\frac{\mathrm{d}c_{ij}^{\mathrm{D}}(t)}{\mathrm{d}t} = W_{ij}^{\text{I-D}}(t) + W_{ij}^{\text{T-D}}(t) - M_{ij}^{\mathrm{D}}(t) - R_{ij}^{\mathrm{D}}(t) \qquad \forall i \in \mathbb{R} \text{ and } j \in \mathbb{U}_i. \tag{3.4}$$

The change in the number of the transferred vehicles at time t is determined by Equations 3.5 and 3.6, where $M_{ii}^{\mathrm{T}}(t)$ and $M_{ij}^{\mathrm{T}}(t)$ denote internal and inter-regional flows of transferred vehicles, respectively. $W_{ij}^{\text{I-T}}(t)$ is the rate of the idle vehicle that are transferred by transferred controller.

$$\frac{\mathrm{d}c_{ii}^{\mathrm{T}}(t)}{\mathrm{d}t} = \sum_{j \in \mathbb{U}_i} M_{ji}^{\mathrm{T}}(t) - M_{ii}^{\mathrm{T}}(t) - W_{ii}^{\text{T-D}}(t) \qquad \forall i \in \mathbb{R}, \tag{3.5}$$

$$\frac{\mathrm{d}c_{ij}^{\mathrm{T}}(t)}{\mathrm{d}t} = W_{ij}^{\text{I-T}}(t) - M_{ij}^{\mathrm{T}}(t) - W_{ij}^{\text{T-D}}(t) \qquad \forall i \in \mathbb{R} \text{ and } j \in \mathbb{U}_i. \tag{3.6}$$

The change in the number of the idle vehicles in each region, $c_i^{\mathrm{I}}(t)$, can be obtained by Equation 3.7. The exogenous arrival and leaving rate of the idle vehicles are denoted by $q_i^{c^+}(t)$ and $q_i^{c^-}(t)$ respectively. $w_i^{\mathrm{I}}(t)$ is the rate of idle vehicles that are close to the boundary of the regions and might leave their current region because of random cruising behavior of idle vehicles.

$$\begin{aligned}\frac{\mathrm{d}c_i^{\mathrm{I}}(t)}{\mathrm{d}t} =& q_i^{c^+}(t) + M_{ii}^{\mathrm{O}}(t) + M_{ii}^{\mathrm{T}}(t) - q_i^{c^-}(t) - \sum_{j\in\{\mathbb{U}_i,i\}} W_{ij}^{\text{I-D}}(t) \\ &- \sum_{j\in\mathbb{U}_i} W_{ij}^{\text{I-T}}(t) + w_i^{\mathrm{I}}(t) \qquad \forall i \in \mathbb{R}.\end{aligned} \tag{3.7}$$

The evolution of the number of waiting passengers are:

$$\frac{\mathrm{d}p_{ii}^{\mathrm{W}}(t)}{\mathrm{d}t} = q_{ii}^{p^{W+}}(t) - q_{ii}^{p^{W-}}(t) - W_{ii}^{\text{W-A}}(t) \qquad \forall i \in \mathbb{R}, \tag{3.8}$$

$$\frac{\mathrm{d}p_{ij}^{\mathrm{W}}(t)}{\mathrm{d}t} = q_{ij}^{p^{W+}}(t) - q_{ij}^{p^{W-}}(t) - W_{ij}^{\text{W-A}}(t) \qquad \forall i \in \mathbb{R} \text{ and } j \in \mathbb{U}_i, \tag{3.9}$$

where, $W_{ii}^{\text{W-A}}(t)$ and $W_{ij}^{\text{W-A}}(t)$ are the rates of the waiting passengers that become assigned at time t. The exogenous arrival rates of the waiting passengers are $q_{ii}^{p^{W+}}(t)$ and $q_{ij}^{p^{W+}}(t)$. $q_{ij}^{p^{W-}}(t)$ and $q_{ij}^{p^{W-}}(t)$ denote the exogenous leaving rate of the waiting passengers.

The number of the assigned passengers can be tracked by Equations 3.10 and 3.11. In these equations, $q_{ii}^{p^{A-}}(t)$ and $q_{ij}^{p^{A-}}(t)$ are the rates of order cancellation of impatient waiting passengers.

$$\frac{\mathrm{d}p_{ii}^{\mathrm{A}}(t)}{\mathrm{d}t} = W_{ii}^{\text{W-A}}(t) - q_{ii}^{p^{A-}}(t) - b_{ii}(t) \qquad \forall i \in \mathbb{R}, \tag{3.10}$$

$$\frac{\mathrm{d}p_{ij}^{\mathrm{A}}(t)}{\mathrm{d}t} = W_{ij}^{\text{W-A}}(t) - q_{ij}^{p^{A-}}(t) - b_{ij}(t) \qquad \forall i \in \mathbb{R} \text{ and } j \in \mathbb{U}_i. \tag{3.11}$$

3.2.2.2 State-Space Formulation

In this subsection, we introduce the mathematical formulation of the model in a standard state-space form to design the transfer controller. The controller for repositioning the vehicles is designed based on the developed model in Chapter 2 that assumes the network is partitioned into a set of regions denoted by $\mathbb{R}$. In the developed model, there are four groups for the vehicles (i.e. idle, transferred, dispatched, and occupied) and three groups for the passengers (i.e. waiting, assigned, and on-board). The state variables that correspond to each group of vehicles and passengers are:

$$\boldsymbol{X}(t)=[\boldsymbol{C}^{\mathrm{I}}(t),\boldsymbol{C}^{\mathrm{T}}(t),\boldsymbol{C}^{\mathrm{D}}(t),\boldsymbol{C}^{\mathrm{O}}(t),\boldsymbol{P}^{\mathrm{W}}(t),\boldsymbol{P}^{\mathrm{A}}(t)]^{\top},\ \boldsymbol{X}(t)\subset\mathcal{R}^{6N+5\sum_{i=1}^{N}\|\mathbb{U}_i\|} \tag{3.12}$$

where $\top$ is the transpose operator and $\mathcal{R}$ denotes the set of real numbers. $\mathbb{U}_i$ is the set of the regions directly connected to Region i and $\|\mathbb{U}_i\|$ is the cardinality of $\mathbb{U}_i$. The number of the regions in the network in denoted by N. Each state of the ride-sourcing system can be decomposed by row-wise concatenation of submatrices such as $c^{\mathrm{T}}_{ik_i}(t)$, $c^{\mathrm{D}}_{ik_i}(t)$, $c^{\mathrm{O}}_{ik_i}(t)$, $c^{\mathrm{D}}_{ik_i}(t)$, $p^{\mathrm{W}}_{ik_i}(t)$, and $p^{\mathrm{A}}_{ik_i}(t)\subset\mathcal{R}^{\left(1+\|\mathbb{U}_i\|\right)\times 1}$, where $\mathcal{R}^{\left(1+\|\mathbb{U}_i\|\right)\times 1}$ denotes the set of real-valued matrices with the size of $\left(1+\|\mathbb{U}_i\|\right)\times 1$. Decomposition of each state is:

$$\boldsymbol{C}^{\mathrm{T}}(t)=\begin{bmatrix}\left[c^{\mathrm{T}}_{1k_1}(t)\right]_{\left(1+\|\mathbb{U}_1\|\right)\times 1}\\ \vdots\\ \left[c^{\mathrm{T}}_{ik_i}(t)\right]_{\left(1+\|\mathbb{U}_i\|\right)\times 1}\\ \vdots\\ \left[c^{\mathrm{T}}_{Nk_N}(t)\right]_{\left(1+\|\mathbb{U}_N\|\right)\times 1}\end{bmatrix},\ \boldsymbol{C}^{\mathrm{D}}(t)=\begin{bmatrix}\left[c^{\mathrm{D}}_{1k_1}(t)\right]_{\left(1+\|\mathbb{U}_1\|\right)\times 1}\\ \vdots\\ \left[c^{\mathrm{D}}_{ik_i}(t)\right]_{\left(1+\|\mathbb{U}_i\|\right)\times 1}\\ \vdots\\ \left[c^{\mathrm{D}}_{Nk_N}(t)\right]_{\left(1+\|\mathbb{U}_N\|\right)\times 1}\end{bmatrix},\ \boldsymbol{C}^{\mathrm{O}}(t)=\begin{bmatrix}\left[c^{\mathrm{O}}_{1k_1}(t)\right]_{\left(1+\|\mathbb{U}_1\|\right)\times 1}\\ \vdots\\ \left[c^{\mathrm{O}}_{ik_i}(t)\right]_{\left(1+\|\mathbb{U}_i\|\right)\times 1}\\ \vdots\\ \left[c^{\mathrm{O}}_{Nk_N}(t)\right]_{\left(1+\|\mathbb{U}_N\|\right)\times 1}\end{bmatrix},$$

$$\boldsymbol{C}^{\mathrm{I}}(t)=\begin{bmatrix}c^{\mathrm{I}}_1(t)\\ \vdots\\ c^{\mathrm{I}}_i(t)\\ \\ \vdots\\ c^{\mathrm{I}}_N(t)\end{bmatrix},\ \boldsymbol{P}^{\mathrm{W}}(t)=\begin{bmatrix}\left[p^{\mathrm{W}}_{1k_1}(t)\right]_{\left(1+\|\mathbb{U}_1\|\right)\times 1}\\ \vdots\\ \left[p^{\mathrm{W}}_{ik_i}(t)\right]_{\left(1+\|\mathbb{U}_i\|\right)\times 1}\\ \vdots\\ \left[p^{\mathrm{W}}_{Nk_N}(t)\right]_{\left(1+\|\mathbb{U}_N\|\right)\times 1}\end{bmatrix},\ \boldsymbol{P}^{\mathrm{A}}(t)=\begin{bmatrix}\left[p^{\mathrm{A}}_{1k_1}(t)\right]_{\left(1+\|\mathbb{U}_1\|\right)\times 1}\\ \vdots\\ \left[p^{\mathrm{A}}_{ik_i}(t)\right]_{\left(1+\|\mathbb{U}_i\|\right)\times 1}\\ \vdots\\ \left[p^{\mathrm{A}}_{Nk_N}(t)\right]_{\left(1+\|\mathbb{U}_N\|\right)\times 1}\end{bmatrix},\tag{3.13}$$

where, $k_i \in \{\mathbb{U}_i, i\}$. $c^{\mathrm{T}}_{ik_i}(t)$, $c^{\mathrm{D}}_{ik_i}(t)$, $c^{\mathrm{O}}_{ik_i}(t)$, $p^{\mathrm{W}}_{ik_i}(t)$, and $p^{\mathrm{A}}_{ik_i}(t)$ are the numbers of the transfer vehicles, dispatched vehicles, occupied vehicles, waiting passengers, and assigned passengers in Region i with final destination in Region k_i at time t. $c^{\mathrm{I}}_i(t)$ denotes the number of the idle vehicles in Region i at time t. We do not track the number of on-board passengers as a state variable because the number of on-board passengers and occupied vehicles are equal at all times.

To obtain the nonlinear state-space form of the model for designing the controller, we categorize the inputs of the ride-sourcing system and the transfer controller into four types:

(i) *Exogenous inputs*: Arrival Rate of the new ride-sourcing vehicles (i.e. drivers who start providing rides), $q^{c^+}_i(t)$, and the rate of idle vehicles that leave the ride-sourcing system (i.e. drivers who finish providing rides), $q^{c^-}_i(t)$, are the external inputs of idle vehicles in Region i. The external demand of the waiting passengers in Region i with destination in Region j, $q^{pW+}_{ij}(t)$, is the rate of the waiting passengers that enter the ride-sourcing system (see Equations 22 to 24 in Chapter 2). The vector of the

exogenous inputs is:

$$
\begin{aligned}
&\boldsymbol{U}^{\text{ex}}(t) = [\boldsymbol{Q}^{c^+}(t), \boldsymbol{Q}^{c^-}(t), \boldsymbol{Q}^{p^{W+}}(t)] \qquad \boldsymbol{U}^{\text{ex}}(t) \subset \mathcal{R}^{3N+\sum_{i=1}^{N} \|\mathbb{U}_i\|},\\
&\boldsymbol{Q}^{\text{c}^+}(t) = [q_1^{\text{c}^+}(t), \ldots, q_i^{\text{c}^+}(t), \ldots, q_N^{\text{c}^+}(t)],\\
&\boldsymbol{Q}^{\text{c}^-}(t) = [q_1^{\text{c}^-}(t), \ldots, q_i^{\text{c}^-}(t), \ldots, q_N^{\text{c}^-}(t)],\\
&\boldsymbol{Q}^{p^{W+}}(t) = \Big[\big[q_{1k_1}^{p^{W+}}(t)\big]_{\left(1+\|\mathbb{U}_1\|\right)\times 1}, \ldots, \big[q_{ik_i}^{p^{W+}}(t)\big]_{\left(1+\|\mathbb{U}_i\|\right)\times 1}, \ldots, \big[q_{Nk_N}^{p^{W+}}(t)\big]_{\left(1+\|\mathbb{U}_N\|\right)\times 1} \Big].
\end{aligned}
\tag{3.14}
$$

(ii) *Endogenous inputs*: these inputs are manipulated by the dispatching subsystem. They are results of the proposed adaptive spatio-temporal matching method in Chapter 2. $W_{ij}^{\text{I-D}}(t)$, $W_{ij}^{\text{T-D}}(t)$, and $W_{ij}^{\text{W-A}}(t)$, $\forall i$ and $j \in \mathbb{R}$, denote the endogenous inputs. $W_{ij}^{\text{I-D}}(t)$ and $W_{ij}^{\text{T-D}}(t)$ are the rates of the idle and transferred vehicles in Region i that are dispatched to Region j respectively. The rate of the waiting passengers in Region i with destination in Region j that are assigned to the vehicles by the matching method is denoted by $W_{ij}^{\text{W-A}}(t)$ (see Equations 3.3 to 3.11). The endogenous inputs are vectorized as:

$$
\begin{aligned}
&\boldsymbol{U}^{\text{en}}(t) = \big[\boldsymbol{W}^{\text{I-D}}(t), \boldsymbol{W}^{\text{T-D}}(t), \boldsymbol{W}^{\text{W-A}}(t)\big] \qquad \boldsymbol{U}^{\text{en}}(t) \subset \mathcal{R}^{3(N+\sum_{i=1}^{N} \|\mathbb{U}_i\|)},\\
&\boldsymbol{W}^{\text{I-D}}(t) = \Big[\big[W_{1k_1}^{\text{I-D}}(t)\big]_{\left(1+\|\mathbb{U}_1\|\right)\times 1}, \ldots, \big[W_{ik_i}^{\text{I-D}}(t)\big]_{\left(1+\|\mathbb{U}_i\|\right)\times 1}, \ldots, \big[W_{Nk_N}^{\text{I-D}}(t)\big]_{\left(1+\|\mathbb{U}_N\|\right)\times 1} \Big],\\
&\boldsymbol{W}^{\text{T-D}}(t) = \Big[\big[W_{1k_1}^{\text{T-D}}(t)\big]_{\left(1+\|\mathbb{U}_1\|\right)\times 1}, \ldots, \big[W_{ik_i}^{\text{T-D}}(t)\big]_{\left(1+\|\mathbb{U}_i\|\right)\times 1}, \ldots, \big[W_{Nk_N}^{\text{T-D}}(t)\big]_{\left(1+\|\mathbb{U}_N\|\right)\times 1} \Big],\\
&\boldsymbol{W}^{\text{W-A}}(t) = \Big[\big[W_{1k_1}^{\text{W-A}}(t)\big]_{\left(1+\|\mathbb{U}_1\|\right)\times 1}, \ldots, \big[W_{ik_i}^{\text{W-A}}(t)\big]_{\left(1+\|\mathbb{U}_i\|\right)\times 1}, \ldots, \big[W_{Nk_N}^{\text{W-A}}(t)\big]_{\left(1+\|\mathbb{U}_N\|\right)\times 1} \Big].
\end{aligned}
\tag{3.15}
$$

The dynamics of the endogenous inputs and how they react to the vehicle repositioning is crucial to determine the optimum rate of idle vehicles transferring. These inputs are considered known in Chapter 2 that reduces the controllability of the model. In Subsection 3.3.2, estimations of the predicted values of the endogenous inputs are presented to capture the dependencies of these inputs and the model.

(iii) *Control inputs*: these inputs are manipulated by the transfer subsystem for dynamically balancing the number of idle vehicles in each region. $W_{ij}^{\text{I-T}}(t)$, $\forall i \in \mathbb{R}$ and $j \in \mathbb{U}_i$, is the control input (see Equations 3.5 and 3.6). This is the rate of the idle vehicles

in Region i that are advised to transfer to Region j by the controller. The control inputs, $\boldsymbol{U}^{\text{co}}(t) \subset \mathcal{R}^{\sum_{i=1}^{N} \|\mathbb{U}_i\|}$, are:

$$\boldsymbol{U}^{\text{co}}(t) = \Big[\big[W^{\text{I-T}}_{1k_1}(t)\big]_{\|\mathbb{U}_1\|\times 1}, \dots, \big[W^{\text{I-T}}_{ik_i}(t)\big]_{\|\mathbb{U}_i\|\times 1}, \dots, \big[W^{\text{I-T}}_{Nk_N}(t)\big]_{\|\mathbb{U}_N\|\times 1} \Big]. \tag{3.16}$$

(iv) *Pseudo inputs*: $R^{\text{D}}_{ij}(t)$ and $q^{p^{W-}}_{ij}(t)$, $\forall i \in \mathbb{R}$ and $j \in \mathbb{U}_i$, are the pseudo inputs in the developed model. $R^{\text{D}}_{ij}(t)$ is the rate of the cancellation of dispatched trips in Region i with pickup location in Region j (see Equations 3.3 and 3.4). $q^{p^{W-}}_{ij}(t)$ is the cancellation rate of the waiting passengers in Region i with destination in Region j leaving the network at time t because of their dissatisfaction from the ride-sourcing system (see Equations 3.8 and 3.9). The value of these memory-less variables are affected by the vehicle-passenger matching method at each time instant. These inputs cannot be directly manipulated by dispatching and transferring subsystems and they are not independent from the state variable. The pseudo inputs are:

$$\begin{aligned}
&\boldsymbol{U}^{\text{ps}}(t) = [\boldsymbol{Q}^{p^{W-}}(t), \boldsymbol{R}^{\text{D}}(t)] \qquad \boldsymbol{U}^{\text{ps}}(t) \subset \mathcal{R}^{2(N+\sum_{i=1}^{N} \|\mathbb{U}_i\|)}, \\
&\boldsymbol{Q}^{p^{W-}}(t) = \Big[\big[q^{p^{W-}}_{1k_1}(t)\big]_{\left(1+\|\mathbb{U}_1\|\right)\times 1}, \dots, \big[q^{p^{W-}}_{ik_i}(t)\big]_{\left(1+\|\mathbb{U}_i\|\right)\times 1}, \dots, \big[q^{p^{W-}}_{Nk_N}(t)\big]_{\left(1+\|\mathbb{U}_N\|\right)\times 1} \Big], \\
&\boldsymbol{R}^{\text{D}}(t) = \Big[\big[R^{\text{D}}_{1k_1}(t)\big]_{\left(1+\|\mathbb{U}_1\|\right)\times 1}, \dots, \big[R^{\text{D}}_{ik_i}(t)\big]_{\left(1+\|\mathbb{U}_i\|\right)\times 1}, \dots, \big[R^{\text{D}}_{Nk_N}(t)\big]_{\left(1+\|\mathbb{U}_N\|\right)\times 1} \Big].
\end{aligned} \tag{3.17}$$

After defining the state variables and the inputs of the system, we can formulate the nonlinear state-space form of the model as

$$\frac{\text{d}\boldsymbol{X}(t)}{\text{d}t} = f\Big(\boldsymbol{X}(t), \boldsymbol{U}^{\text{ex}}(t), \boldsymbol{U}^{\text{en}}(t), \boldsymbol{U}^{\text{co}}(t), \boldsymbol{U}^{\text{ps}}(t)\Big), \tag{3.18}$$

where, $f(.) : \mathcal{R}^{6N+5\sum_{i=1}^{N} \|\mathbb{U}_i\|} \times \mathcal{R}^{3N+\sum_{i=1}^{N} \|\mathbb{U}_i\|} \times \mathcal{R}^{3(N+\sum_{i=1}^{N} \|\mathbb{U}_i\|)} \times \mathcal{R}^{\sum_{i=1}^{N} \|\mathbb{U}_i\|} \times \mathcal{R}^{2(N+\sum_{i=1}^{N} \|\mathbb{U}_i\|)} \rightarrow \mathcal{R}^{6N+5\sum_{i=1}^{N} \|\mathbb{U}_i\|}$ is a continuous differentiable map (see Equations 16 to 26 in Chapter 2). The transferring subsystem manipulates $\boldsymbol{U}^{\text{co}}(t)$ to regulates the state variables to satisfy a defined objective function. $\boldsymbol{U}^{\text{co}}(t) = 0$ reflects a ride-sourcing system without transferring subsystem. In the next section, we introduce the NMPC for regulating the state variables.

3.3 Nonlinear Model Predictive Controller Design

In this section, we introduce the NMPC as the repositioning controller of the proposed ride-sourcing system. The interconnection between Dispatching Subsystem, Plant, and the NMPC is illustrated in Figure 3.1. Model predictive controllers optimize the future behavior of the system. To this end, the controller is required to predict the state variables of the system. The prediction is determined based on the validated developed model (see Chapter 2) and its corresponding nonlinear state-space representation (see Equation 3.18). The nonlinear state-space model for predicting the state variables needs the pseudo inputs, estimated values of the exogenous inputs, $\widehat{\boldsymbol{U}}^{\text{ex}}(t)$, and estimated values of endogenous inputs $\widehat{\boldsymbol{U}}^{\text{en}}(t)$. Estimated values of exogenous inputs are determined by applying exponential moving average on previous values of the exogenous inputs to clean up the noisy values. To increase the controllability of the model, the estimated values of the k steps ahead predicted endogenous inputs, $\widehat{\boldsymbol{U}}^{\text{en}}(t+k)$, are determined as a function of the predicted state variables, $\boldsymbol{X}(t+k); k>0$ (see Subsection 3.3.2).

The predicted values of the state variables and constraints that are defined based on the nature of the system is fed to the optimization problem to minimize the defined cost function (see Equation 3.19). The NMPC does not have a direct connection with Dispatching Subsystem; hence, the designed NMPC can be integrated with any Dispatching Subsystems. The *Link-Level Allocation* obtains the output of the NMPC that is the inter-*regional* desired rates of the transferred vehicles to recommend hot-spot *streets* to reposition idle vehicles.

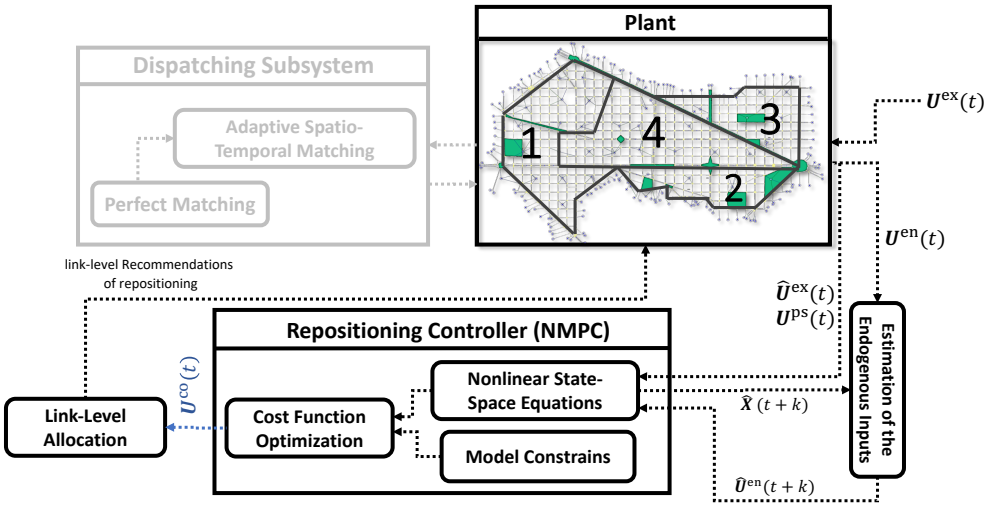


FIGURE 3.1. Interaction between the Repositioning Controller and the Plant. (Note that the focus of this Chapter is not the Dispatching Subsystem and hence it is illustrated with lighter shades.) The exogenous demand of the trip requests and incoming vehicles are generated in the traffic network defined by $\boldsymbol{U}^{\text{ex}}(t)$. The estimation of the predicted vehicle-passenger matchings, $\widehat{\boldsymbol{U}}^{\text{en}}(t+k)$, the cancellation rate of the dispatched vehicles and passengers' trip requests, $\boldsymbol{U}^{\text{ps}}(t)$, and estimated values of the exogenous inputs, $\widehat{\boldsymbol{U}}^{\text{ex}}(t)$, are utilised in Nonlinear State-Space Equations to predict the state variables, $\boldsymbol{X}(t+k)$. The set of controller inputs, $\boldsymbol{U}^{\text{co}}(t)$, are determined by solving the constrained optimization problem.

In the following, first, we define the cost function and constraints of the controller. In the second part, we propose estimations of the endogenous inputs during prediction horizon to improve the controllability and pro-activeness of the controller.

3.3.1 Cost Function and Constraints

The aim of the transfer subsystem is to reposition idle vehicles to the regions with an excess number of passengers to increase possibility of successful matchings in the future. The NMPC solves at every transferring time instance the finite-horizon optimal control problem. NMPC solves Equation 3.19 to control the rate of the transferring vehicles, $\boldsymbol{U}^{\text{co}}(t)$, to minimize total *predicted* waiting time of the vehicles (i.e. the difference between drop-off time and the next matching time) and total *predicted* waiting time of the passengers (i.e. the time that passengers

remain unmatched in the network).

$$\min_{\boldsymbol{U}^{\mathrm{co}}(t),\ldots,\boldsymbol{U}^{\mathrm{co}}(t+(n_{\mathrm{c}}-1)\times\Delta t_{\mathrm{pr}})} J\big(\boldsymbol{X}(t),\boldsymbol{U}^{\mathrm{ex}}(t),\boldsymbol{U}^{\mathrm{en}}(t),\boldsymbol{U}^{\mathrm{co}}(t)\big)$$
$$= \min_{\boldsymbol{U}^{\mathrm{co}}(t_{\mathrm{tr}}),\ldots,\boldsymbol{U}^{\mathrm{co}}(t_{\mathrm{tr}}+(n_{\mathrm{c}}-1)\times\Delta t_{\mathrm{pr}})} \int_{t_{\mathrm{tr}}}^{t_{\mathrm{tr}}+n_{\mathrm{p}}\times\Delta t_{\mathrm{pr}}} \big[\boldsymbol{C}^{\mathrm{I}}(t),\boldsymbol{C}^{\mathrm{T}}(t),\boldsymbol{P}^{\mathrm{W}}(t)\big]\boldsymbol{Q}\big[\boldsymbol{C}^{\mathrm{I}}(t),\boldsymbol{C}^{\mathrm{T}}(t),\boldsymbol{P}^{\mathrm{W}}(t)\big]^{\top}\mathrm{d}t, \tag{3.19}$$

where $\boldsymbol{Q} = \mathrm{diag}\big(\boldsymbol{Q}^{\mathrm{I}},\boldsymbol{Q}^{\mathrm{T}},\boldsymbol{Q}^{\mathrm{W}}\big)$. $\boldsymbol{Q}^{\mathrm{I}} \subset \mathcal{R}^{N}$, $\boldsymbol{Q}^{\mathrm{T}}$, and $\boldsymbol{Q}^{\mathrm{W}} \subset \mathcal{R}^{N+\sum_{i=1}^{N}\|\mathbb{U}_i\|}$ are symmetric positive semidefinite matrices for weightening of contributions of idle vehicles, transfer vehicles, and waiting passengers in the cost function respectively. These matrices can be used to regulate the significance of vehicles (drivers) and passengers for the ride-sourcing company. $\int \boldsymbol{C}^{\mathrm{I}}(t)\boldsymbol{Q}^{\mathrm{I}}\boldsymbol{C}^{\mathrm{I}^{\top}}(t) + \boldsymbol{C}^{\mathrm{T}}(t)\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}^{\top}}(t)\mathrm{d}t$ formulates the weighted square of the total predicted waiting time of the vehicles. The total predicted waiting time of the passengers is denoted by $\int \boldsymbol{P}^{\mathrm{W}}(t)\boldsymbol{Q}^{\mathrm{W}}\boldsymbol{P}^{\mathrm{W}^{\top}}(t)\mathrm{d}t$. t_{tr} is the transfer time instance and $n_{\mathrm{p}}\times\Delta t_{\mathrm{pr}}$ is the prediction duration. Δt_{pr} denotes the prediction sample and n_{p} is the prediction horizon. We can utilize the control horizon of n_{c} time steps to optimize only the first n_{c} steps of the transfer actions ($n_{\mathrm{c}} \leq n_{\mathrm{p}}$). This reduces the computational efforts for solving the optimization problem and the feasibility of the problem.

At each transfer time instance, the NMPC uses feedback from the ride-sourcing system to obtain the state variables, $\boldsymbol{X}(t_{\mathrm{tr}})$, exogenous inputs, $\boldsymbol{U}^{\mathrm{ex}}(t_{\mathrm{tr}})$, and endogenous inputs, $\boldsymbol{U}^{\mathrm{en}}(t_{\mathrm{tr}})$. NMPC uses these values as the initial value of the prediction and utilize the developed model (see Equation 3.18) to make prediction of the state of the system up to time $t_{\mathrm{tr}}+n_{\mathrm{p}}\times\Delta t_{\mathrm{pr}}$. It solves the optimization problem to find the sequence of optimum transfer actions, $\big[\boldsymbol{U}^{\mathrm{co}^*}(t_{\mathrm{tr}}),\boldsymbol{U}^{\mathrm{co}^*}(t_{\mathrm{tr}}+\Delta t_{\mathrm{pr}}),\ldots,\boldsymbol{U}^{\mathrm{co}^*}(t_{\mathrm{tr}}+(n_{\mathrm{c}}-1)\times\Delta t_{\mathrm{pr}})\big]$, to minimize the total predicted waiting time. The solution of the optimization problem must satisfy Constraints 3.20. Afterward, only the first step of the obtained optimum transfer actions is applied for transferring idle vehicles,

$\boldsymbol{U}^{\mathrm{co}^*}(t_{\mathrm{tr}})$.

$$\boldsymbol{X}(t) \geq 0 \qquad t \in [t_{\mathrm{tr}}, t_{\mathrm{tr}}+n_{\mathrm{p}}\times\Delta t_{\mathrm{pr}}], \tag{3.20a}$$

$$\boldsymbol{U}^{\mathrm{co}}(t) \geq 0 \qquad t \in [t_{\mathrm{tr}}, t_{\mathrm{tr}}+n_{\mathrm{p}}\times\Delta t_{\mathrm{pr}}], \tag{3.20b}$$

$$\sum_{j\in\{\mathbb{U}_i,i\}} (c_{ij}^{\mathrm{T}}(t)+c_{ij}^{\mathrm{D}}(t)+c_{ij}^{\mathrm{O}}(t)) \leq n_i^{\mathrm{jam}}-n_i(t_{\mathrm{tr}})-c_i^{\mathrm{I}}(t) \quad t \in [t_{\mathrm{tr}}, t_{\mathrm{tr}}+n_{\mathrm{p}}\times\Delta t_{\mathrm{pr}}], \forall i \in \mathbb{R}, \tag{3.20c}$$

where, $n_i(t_{\mathrm{tr}})$ is the total number of the non-ride-sourcing vehicles in Region i at time t_{tr}. Since the model does not track the evolution of the number of non-ride-sourcing vehicles, we assume that this number remains fix during the prediction horizon. In each transferring time, this number is updated by feedback from the system. Constraints 3.20a and 3.20b ensure the positivity of all the state variables and optimum transferring rates. Constraint 3.20c sets an upper limit on total number of the vehicles in each region. An upper-bound for each state variables can be set easily to enforce the optimum transferring rates that are aligned with policies of the ride-sourcing company. For more details about MPC and its use in ride-sourcing systems and traffic control see for example (Ramezani and Nourinejad 2018b; Yildirimoglu et al. 2018).

Optimization problem of Equation 3.19 is solved by using interior method for nonlinear programming that combines line search and trust region steps. In this method, the primary step is line search step using direct linear algebra. If the problem is not convex or there is singularity in Jacobian and Hessian then the primary step will be replaced by trust region step. This replacing of trust region step guarantees progress of the optimization when the line search step fails. In this method, Karush–Kuhn–Tucker (KKT) approach is used to consider the equality and inequality constraints (Waltz et al. 2006). This solver is available in KNITRO (Byrd et al. 2006) and MATLAB and computational efficiency of this solver is well-studied in (Wächter and Biegler 2006).

3.3.2 Estimation of the Endogenous Inputs

We consider three sampling times in the system: matching, transfer, and prediction clocks. As depicted in Figure 3.2, the matching frequency, Δt_{m_i}, is time-varying. The frequency of the matching in this Chapter is determined from the proposed adaptive spatio-temporal matching method in Chapter 2. In this study, we assume the frequency of transferring is time-invariant. To set the transferring frequency, taking into account the average number of rejection that is determined from the adaptive spatio-temporal filtering is crucial. The higher the number of rejections shows the higher necessity of transferring execution. This is to have sufficient imbalanced idle vehicles available in the network. The prediction frequency is introduced to discretize the continuous model (Equation 3.18) for the prediction of the state variables. At each transfer time, t_{tr}, we assume prediction sampling time, Δt_{pr}, is time-invariant.

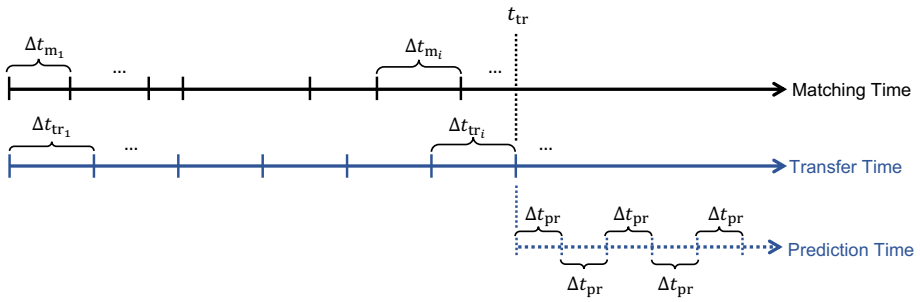


FIGURE 3.2. Comparison between matching, transfer, and prediction sample times.

The transfer subsystems is executed at time t_{tr}. The endogenous inputs (e.g. the rate of the idle/transferred vehicles to be matched to the waiting passengers) and pseudo inputs are known at any time before $t \leq t_{\mathrm{tr}}$. The pseudo inputs are not dominant inputs to change the operating point of the system and intrinsically are stochastic processes because they reflect human preferences. Hence, we consider them as disturbances in designing NMPC. The endogenous inputs that are dominant factors are *not known* during the prediction time, $t \in (t_{\mathrm{tr}}, t_{\mathrm{tr}} + n_{\mathrm{p}} \times \Delta t_{\mathrm{pr}}]$. In addition, the evolution of the state variables of the model is a function of the endogenous inputs and availability of them is necessary for prediction of the state variables. Our experiments show that by assuming fix values for these inputs during the

prediction interval, the state-space model formulation becomes uncontrollable. The reason is that the effect of manipulating control inputs, $\boldsymbol{U}^{\text{co}}(t)$, cannot be reflected on rate of the vehicles-passenger matching during the prediction interval. To address this issue, let us denote

$$m^{\text{myopic}}(t) = \min\Big(\sum_{i\in\mathbb{R}} c_i^{\text{I}}(t) + \sum_{i\in\mathbb{R}}\sum_{j\in\{\mathbb{U}_i,i\}} c_{ij}^{\text{T}}(t), \sum_{i\in\mathbb{R}}\sum_{j\in\{\mathbb{U}_i,i\}} p_{ij}^{\text{W}}(t)\Big) \quad t\in[t_{\text{tr}}, t_{\text{tr}}+n_{\text{p}}\times\Delta t_{\text{pr}}], \tag{3.21}$$

$$r = \frac{m^{\text{plant}}(t_{\text{tr}})}{m^{\text{myopic}}(t_{\text{tr}})}, \tag{3.22}$$

where, $m^{\text{plant}}(t_{\text{tr}})$ represents the number of the vehicles-passenger matching at time t_{tr} (with considering rejection) and $m^{\text{myopic}}(t)$ denotes the number of the matchings without considering rejection. r that is the rejection ratio is assumed to be constant during the prediction interval.

The endogenous inputs during the prediction interval can be estimated as Equations 3.23-3.25. At each transfer time instance, the NMPC uses the feedback from the system to update $m^{\text{plant}}(t_{\text{tr}})$ and $m^{\text{myopic}}(t_{\text{tr}})$ to update the rejection ratio.

$$\widehat{W}_{ij}^{\text{I-D}}(t) = \frac{\sum_{i\in\{\mathbb{U}_j,i\}} p_{ji}^{\text{W}}(t)}{\sum_{i\in\mathbb{R}}\sum_{j\in\{\mathbb{U}_i,i\}} p_{ij}^{\text{W}}(t)} \frac{c_i^{\text{I}}(t)}{\sum_{i\in\mathbb{R}} c_i^{\text{I}}(t) + \sum_{i\in\mathbb{R}}\sum_{j\in\{\mathbb{U}_i,i\}} c_{ij}^{\text{T}}(t)} \times r \times \frac{m^{\text{myopic}}(t)}{\Delta t_{\text{pr}}}, \tag{3.23}$$

$$\widehat{W}_{ij}^{\text{T-D}}(t) = \frac{\sum_{i\in\{\mathbb{U}_j,i\}} p_{ji}^{\text{W}}(t)}{\sum_{i\in\mathbb{R}}\sum_{j\in\{\mathbb{U}_i,i\}} p_{ij}^{\text{W}}(t)} \frac{\sum_{j\in\{\mathbb{U}_i,i\}} c_{ij}^{\text{T}}(t)}{\sum_{i\in\mathbb{R}} c_i^{\text{I}}(t) + \sum_{i\in\mathbb{R}}\sum_{j\in\{\mathbb{U}_i,i\}} c_{ij}^{\text{T}}(t)} \times r \times \frac{m^{\text{myopic}}(t)}{\Delta t_{\text{pr}}}, \tag{3.24}$$

$$\widehat{W}_{ij}^{\text{W-A}}(t) = \frac{p_{ij}^{\text{W}}(t)}{\sum_{i\in\mathbb{R}}\sum_{j\in\{\mathbb{U}_i,i\}} p_{ij}^{\text{W}}(t)} \times r \times \frac{m^{\text{myopic}}(t)}{\Delta t_{\text{pr}}}, \tag{3.25}$$

where, $t\in[t_{\text{tr}}, t_{\text{tr}}+n_{\text{p}}\times\Delta t_{\text{pr}}]$. $\widehat{W}_{ij}^{\text{I-D}}(t)$ and $\widehat{W}_{ij}^{\text{T-D}}(t)$ respectively denote the estimated rate of the idle and transferred vehicles with current location in Region i that become dispatched in Region j (the origin of the passenger is in Region j) at time t. $\widehat{W}_{ij}^{\text{W-A}}(t)$ is the estimated rate of the waiting passengers with pick up location in Region i and drop off location in Region j at time t who become assigned. The first terms in RHS of Equations 3.23 and 3.24 are the total number of the waiting passengers in Region j that is normalised with respect to total number of the waiting passengers in the network. This ratio reflects the portion of the

dispatched vehicles to Region j. The second terms in RHS of Equations 3.23 and 3.24 reflect the portion of the idle and the transferred vehicles, respectively, that the dispatched vehicles are selected from. The last two terms in RHS of Equations 3.23 to 3.25 estimate the rate of matchings with rejection.

3.4 Microsimulation Experiments

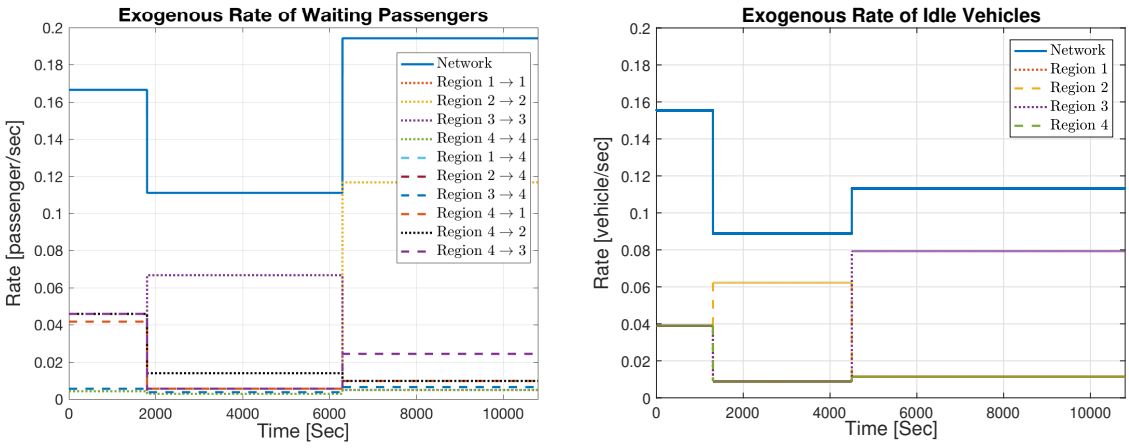
In this section, we introduce the traffic microsimulation case study and illustrate results of applying three scenarios. In the first scenario, there is no transferring subsystem. In the second scenario, we use a reactive controller as in Section 3.4.2. The reactive controller is developed based on transferring the vehicles to the regions with a higher instantaneous number of waiting passengers. The third scenario uses NMPC as the transferring subsystem that considers the effect of the transferring on future evolution of the state variables. In all the scenarios, we utilise the adaptive spatio-temporal matching that is introduced in Chapter 2 as the dispatching subsystem. Hence, the discarded vehicles that is obtained by execution of the matching method and new idle vehicles arrival are considered as potential idle vehicles for transferring to neighbor regions. In these experiments, the only difference between all the scenarios is the transferring subsystem. In the following, we compare the evolution of the idle, transferred, dispatched, and occupied vehicles as well as waiting and assigned passengers for each transferring scenario. Moreover, we compare all the scenarios with respect to passengers' and vehicles' unassigned time, passengers' and vehicles' waiting time, the number of trip cancellations, and the number of the impatient drivers who leave the system because of not being matched to passengers for an excessive time.

3.4.1 Case Study

We scrutinize the efficiency of the proposed transferring controller with a traffic microsimulation. To this end, we develop a ride-sourcing testbed in Aimsun in which ride-sourcing vehicles and travel demands fully interact with other modes of traffic (private vehicles and buses). This microsimulation setup includes point-to-point routing of ride-sourcing vehicles,

assigning vacant vehicles (i.e. transfer and idle vehicles) to waiting passengers, managing generation of emerging ride-sourcing vehicles, and simulating passenger's travel requests. Routing of the vehicles is determined by finding the shortest travel time path between designated origin and destination. In this study, we utilize the adaptive spatio-temporal matching method that is described in Chapter 2 for assigning vacant vehicles to waiting passengers and 60 [sec] and 5 [sec] are set as the upper and lower bounds of this matching method. New idle ride-sourcing vehicles and waiting passengers join the network in a streamline fashion. A stochastic time threshold is considered for all the idle ride-sourcing vehicles and waiting passengers to model impatient characteristic of them. If a driver remains vacant or a waiting passenger is not picked up before their threshold time, they leave the ride-sourcing system or cancel their trip request because of an unsatisfactory quality of service.

To investigate the effectiveness of the NMPC as the transferring subsystem, we apply the proposed controller on the calibrated traffic microsimulation model of part of Barcelona that is partitioned into four regions as in (Kouvelas et al. 2017). To this end, we set the transfer cycle equal to 60 [sec] to assure there are sufficient idle vehicle candidates for transferring. Unbalance supply-demand scenario is created by setting the the exogenous rate of the travel requests (i.e. arrival of waiting passengers) and the exogenous rate of the arrival of new ride-sourcing vehicles as in Figure 3.3. In this figure, during $[1800, 4500]$ and $[6300, 10800]$, most of the travel requests have origins and destinations in Region 3 and Region 2 respectively. While majority of the new idle vehicles appear in Region 2 and Region 3 respectively. Since Region 2 and Region 3 are far from each other (see Figure 3.1), the distance between the majority of idle vehicles and waiting passengers are greater than the maximum matching distance that is obtained from the adaptive spatio-temporal matching method. In each transfer time instance, the exogenous inputs are estimated during the prediction horizon with exponential smoothing. The weight is set to 0.7 and the initial condition of the smoothed value is set to the initial value of the exogenous inputs at the beginning of each transfer time instance.



(A) (B)

FIGURE 3.3. Arrival rate of new (a) waiting passengers and (b) idle vehicles who join the ride-sourcing system.

3.4.2 Reactive Transferring

In this part, we introduce a model-free reactive controller as a benchmark method to scrutinize the pro-activeness of the NMPC. The reactive controller uses the instantaneous number of idle vehicles and waiting passengers at each transferring time to determine the desired number of idle vehicles in each region. Then, it calculates the repositioning rates of idle vehicle between regions to achieve the desired number of idle vehicles in each region. The desired number of idle vehicles in Region i at time t, $c^{*\mathrm{I}}_i(t)$, is obtained by solving the following constrained optimization:

$$\begin{aligned} \min_{c^{*\mathrm{I}}_1(t),\dots,c^{*\mathrm{I}}_N(t)} & \left(\sum_{i\in\mathbb{R}} \left(c^{*\mathrm{I}}_i(t) - \sum_{j\in\{\mathbb{U}_i,i\}} p^{\mathrm{W}}_{ij}(t) \right) \right)^2 \\ \text{s.t.} \quad & \sum_{i\in\mathbb{R}} c^{*\mathrm{I}}_i(t) = \sum_{i\in\mathbb{R}} c^{\mathrm{I}}_i(t). \end{aligned} \tag{3.26}$$

Equation 3.26 determines the number of desired idle vehicles such that the difference between idle vehicles and waiting passengers becomes minimum in all regions. This is to avoid long-distance matchings (i.e. reducing the rejection rate of vehicle-passenger matchings).

The constraint is that the total number of the desired idle vehicles must be equal to the total number of idle vehicles in the ride-sourcing system at time t.

To reach the desired spatial distribution as in Equation 3.26, a subset of idle vehicles in each region are advised to transfer to other regions. To this end, we need to determine the transferring origin (i.e. choosing designated idle vehicle for transferring) and the transferring destination for each of the idle vehicles. Transferring origins and destinations are obtained by solving an integer optimization problem heuristically that is equivalent to finding a subset of a maximum matching of a bipartite graph. In this problem, each side of the bipartite graph is partitioned to a set of disjoint subsets (i.e., supply and demand) (see Figure 3.4). A subset is defined as *supply* region if there is excess number of idle vehicles, $c^{*\mathrm{I}}_i(t) - c^{\mathrm{I}}_i(t) < 0$, and is defined as *demand* region if there is shortage of idle vehicles, $c^{*\mathrm{I}}_i(t) - c^{\mathrm{I}}_i(t) > 0$. If current number of the idle vehicles in each region is equal to the desired number of the idle vehicles, the reactive controller does not send in or send out any idle vehicles.

The bipartite graph as depicted in Figure 3.4, $G(U, V, E)$, has two set of vertices (i.e. supply, U, and demand, V) and one set of edges E. Each edge $e = (u, v) \in E, u \in U, v \in V$ has a weight, $w(e)$. The set of supply vertices, U, is partitioned into κ disjoint subsets that κ is the number of supply regions, $U_1, \ldots, U_\kappa$. The set of demand vertices, V, is partitioned into κ' disjoint subsets that κ' is the number of supply regions, $V_1, \ldots, V_{\kappa'}$.

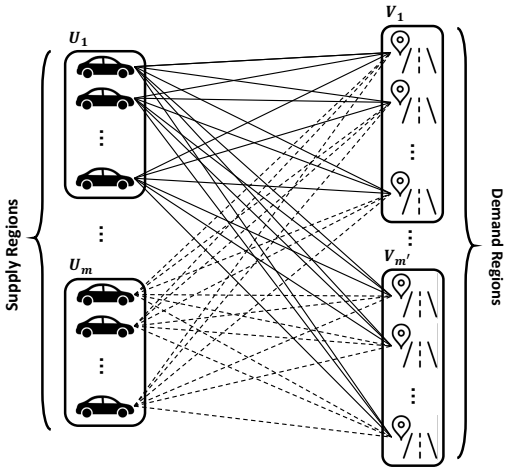


FIGURE 3.4. The bipartite graph for transferring idle vehicles to candidate locations that is obtained from the reactive controller.

The cardinality of each subset is

$$\|U_i\| = c_i^{\mathrm{I}}(t) \quad i \in \{1, \ldots, \kappa\};\ \|V_i\| = c^{*\mathrm{I}}{}_i(t) - c_i^{\mathrm{I}}(t) \quad i \in \{1, \ldots, \kappa'\}. \tag{3.27}$$

Each supply region has $c_i^{\mathrm{I}}(t) - c^{*\mathrm{I}}{}_i(t)$ excess of idle vehicles that must be transferred to locations in demand regions. Each demand region needs $c^{*\mathrm{I}}{}_i(t) - c_i^{\mathrm{I}}(t)$ idle vehicles to be transferred from supply region. Nodes in Figure 3.4 in the demand regions are candidate locations (i.e. destination of transferring path). The candidate locations can be chosen randomly around the geometric center of available waiting passengers of the demand region. Nodes in Figure 3.4 in supply regions are the current location of idle vehicles (i.e. origin of transferring path). The weights, $w(e, v)$, of the graph are negative of shortest travel time between the idle vehicles in supply regions and the candidate locations in the demand regions that are determined by Dijkstra algorithm. The maximum matching in each supply subset, U_i, is $c_i^{\mathrm{I}}(t) - c^{*\mathrm{I}}{}_i(t)$. By solving subsets of maximum matching of G (Barketau et al. 2015), we determine pairs of origins and destinations for transferring idle vehicles at each transferring time.

3.4.3 Evolution of State of the Ride-sourcing System

In this part, the evolution of the number of different groups of vehicles (idle, dispatched, occupied, and transferred) and passengers (waiting and assigned) are depicted for the center region (Region 4) and the surrounding regions (Regions 1, 2, and 3). The center region has borders with all the surrounding regions but the surrounding regions have not borders with each other. We investigate the effect of transferring subsystem by using no transferring, reactive controller, and NMPC. The ride-sourcing system in all scenarios utilises the adaptive spatio-temporal matching method. The frequency of the transfer subsystem is fixed at 60 [sec]. The results of the NMPC is reported by setting $n_{\mathrm{c}} = 2$, $n_{\mathrm{p}} = 30$, and $\Delta t_{\mathrm{pr}} = 15$ [sec]. Choosing a proper value for the prediction and control horizons depends on the dynamic of the system. There is a trad-off in choosing proper value for prediction horizon: choosing a very low value for the prediction horizon results in a reactive controller; however, choosing a very large value for the prediction horizon results in missing regulating the transient behavior

of the system. The control horizon manipulates the complexity of the optimization problem. We set the control horizon equal to 2 to reduce the complexity.

Figure 3.5 shows the evolution of the number of the waiting passengers and idle vehicles by using no transferring, reactive controller, and NMPC. At the start of the test, there is an excess number of idle vehicles and waiting passengers simultaneously that shows the high number of long-distance rejections in the center region (i.e. Region 4). The high number of the long-distance rejections stems from the fact that the majority of the travel requests are generated in the center region but the idle vehicles are distributed uniformly in the center and surrounding regions (see Figure 3.3). Based on Figures 3.5.a, 3.5.b, and 3.5.c, the number of the waiting passengers at the beginning of the tests are the same despite the transferring scenarios. However, the number of the idle vehicles by using a reactive controller or NMPC is smaller than no transferring scenario (see Figures 3.5.d, 3.5.e, and 3.5.f) because some of the idle vehicles become transferred.

During $[0, 1500]$ (sample 0 to 100), the number of the waiting passengers decreases in the center region (i.e. Region 4) in all scenarios because of two reasons (i) more ride-sourcing vehicles join the system and (ii) Region 4 is at the center of the network and has a border with other surrounding regions and vehicles in the surrounding regions can be matched to waiting passengers in the center region with a higher possibility. The rate of the decreasing of the number of the waiting passengers in the center region by using the suggested NMPC is higher than the reactive controller and the no transferring scenarios (see Figures 3.5.a, 3.5.b, and 3.5.b) that is the result of well positioning of the vacant vehicles by the repositioning subsystem.

Figure 3.5 shows that there is an excess number of the waiting passengers in Region 2 and idle vehicles in Region 3 after time 6150 [sec] (sample 410) because (i) there is an abrupt change in passengers' trip requests in Region 2 with a destination in Region 2 (see Figure 3.3.a) and (ii) the rate of exogenous idle vehicles increases drastically in Region 3 after time 4500 [sec] (sample 300) as depicted in Figure 3.3.b. We see that in Figures 3.5.a and 3.5.d after time 6150 [sec] (sample 410), the number of the waiting passengers is mainly higher than the number of the waiting passengers by using reactive controller and NMPC because Region

3 and Region 2 are far from each other and the adaptive spatio-temporal matching method discards majority of vehicle-passenger matchings between the waiting passengers in Regions 2 and the idle vehicles in Region 3. Figures 3.5.b, 3.5.c, 3.5.e, and 3.5.f illustrate that the reactive controller and NMPC shrink the unbalance between waiting passengers and idle vehicles by reducing the number of the waiting passengers in Region 2 and idle vehicles in Region 3. Figure 3.6 show the number of the transferred vehicles in each region at time t. This figure depicts that the excess idle vehicles in Region 3 and Region 4 are transferred to increase the probability of being matched to the waiting passengers in Region 2.

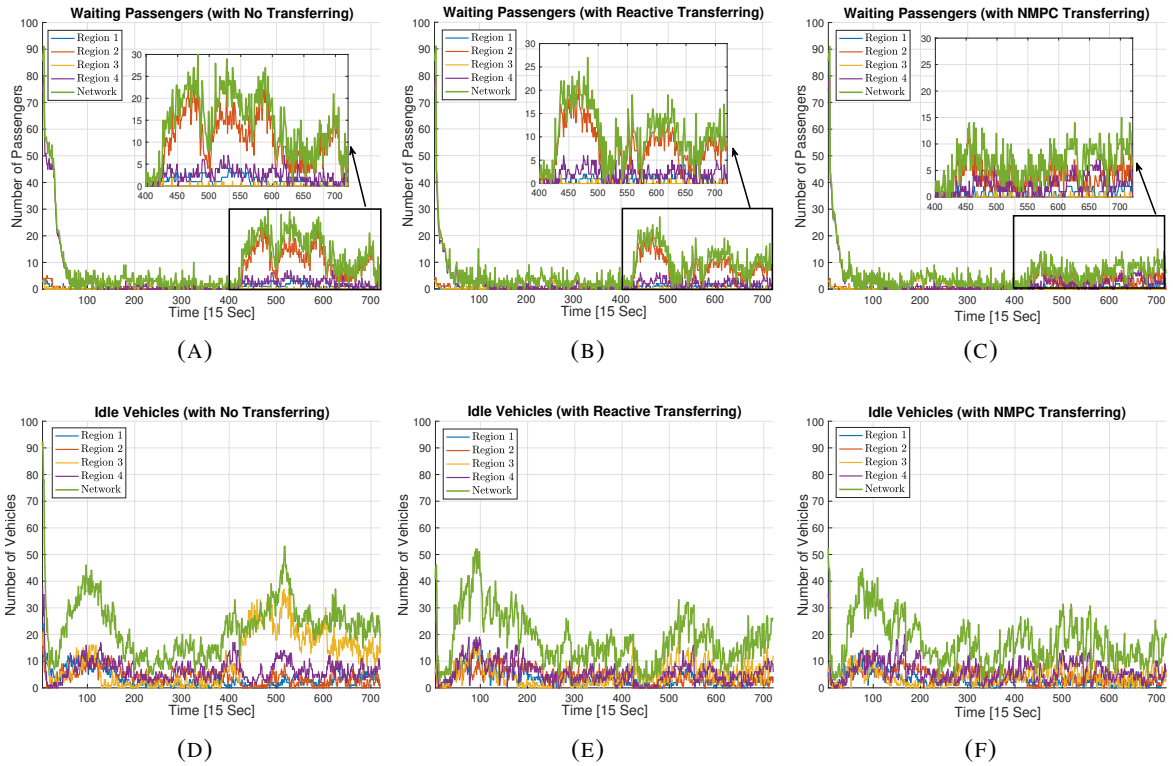
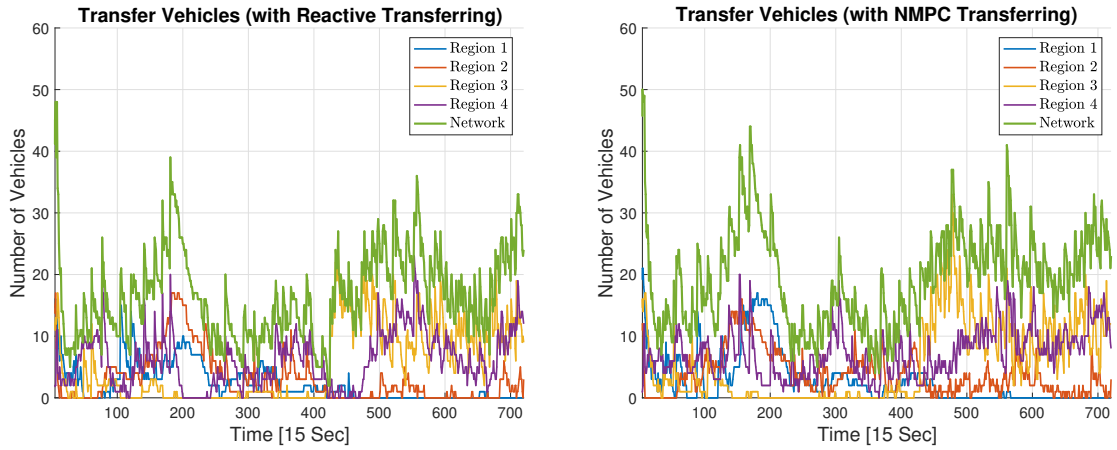


FIGURE 3.5. Evolution of the number of the waiting passengers and idle vehicles by using different transferring methods. Waiting Passengers (a) without transferring, (b) with reactive transferring, and (c) NMPC transferring. Idle vehicles (d) without transferring, (e) with reactive transferring, and (f) NMPC transferring.



(A) (B)

FIGURE 3.6. Evolution of the number of the transferred vehicles by utilizing the (a) reactive controller and (b) NMPC. Each curve represents the number of the transferred vehicles at each region. For example blue curve depicts the number of transferred vehicles in Region 1 independent of their origins and destinations.

Figure 3.7 shows the non-zero transferring rates of the idle vehicles by the NMPC. The total numbers of the repositioning trips with the reactive controller and NMPC are 422 and 526 respectively. The NMPC transfers the vehicles from the surrounding regions (i.e. Regions 1, 2, and 3) to the center region (i.e. Region 4). In this figure, during $[0, 3750]$ (sample 0 to 250), the rate of the vehicles that is transferred from Region 2 to the center region increases because in time $[0, 1800]$ (samples 0 to 120) the exogenous rate of the waiting passengers from Region 4 to Region 2 is relatively high (see Figure 3.3.a) that results in accumulation of the idle vehicles in Regions 2 (ride-sourcing vehicles pickup passengers in Region 4 and drop off them in Region 2 and become idle in Region 2). Furthermore, at time 1200 [sec] (sample 80) the rate of new (exogenous) idle vehicles in Region 2 increases significantly (see Figure 3.3.b) and remains relatively high till 4500 [sec] (sample 300). It amplifies the accumulation of the idle vehicles in Region 2. After sample 460 in Figure 3.6 (time 6900 [sec]), the rate of the transferred vehicles from Region 3 to Region 4 increases to compensate (i) the excess of waiting passengers in Region 2 that is a result of the exogenous rate of the waiting passengers

from Region 2 to Region 2 (see Figure 3.3.a) and (ii) excess of idle vehicles in Region 3 that is a result of the exogenous rate of the idle vehicles in Region 3 (see Figure 3.3.b).

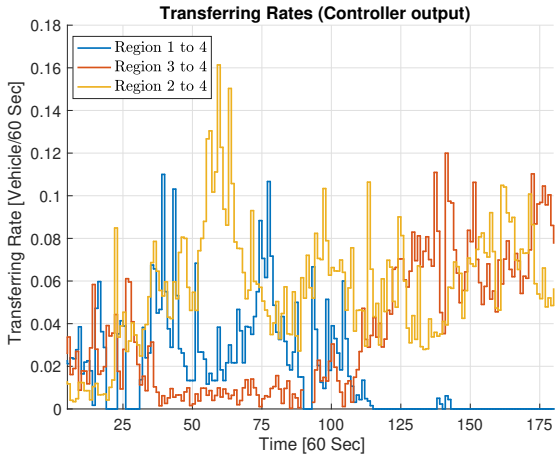


FIGURE 3.7. The non-zero transferring rates implemented by NMPC.

Figure 3.8 illustrates the evolution of the numbers of the dispatched vehicles, occupied vehicles, and assigned passengers in the system with no transferring method, reactive controller, and NMPC. During samples between 0 to 100, there are excess numbers of the waiting passengers and idle vehicles in Region 4 (see Figure 3.5). During this time, the total number of the dispatched vehicles, Figures 3.8.a, 3.8.b, and 3.8.c, (equivalent to the total number of assigned passengers, Figures 3.8.g, 3.8.h, and 3.8.i) of reactive controller is relatively higher than no transferring and NMPC. However, During samples 60 to 200, the total number of the occupied vehicles by using NMPC is relatively higher than no transferring and reactive controller. During this period, the mean values of the occupied vehicles with no transferring, reactive controller, and NMPC are 54.3, 60.2, and 64.3, respectively. It shows NMPC is effective for reducing deadheading time/mileage. Moreover, after sample 400 when the transferring methods reduce the unbalance between idle vehicles in Region 3 and waiting passengers in Region 2, the mean values of the total number of the occupied vehicles with no transferring, reactive controller, and NMPC are 70.4, 73.6, and 80.2 respectively. This is 13.9% and 8.9% improvements in occupation rate of vehicles by using the NMPC in comparison with no transferring and the reactive controller.

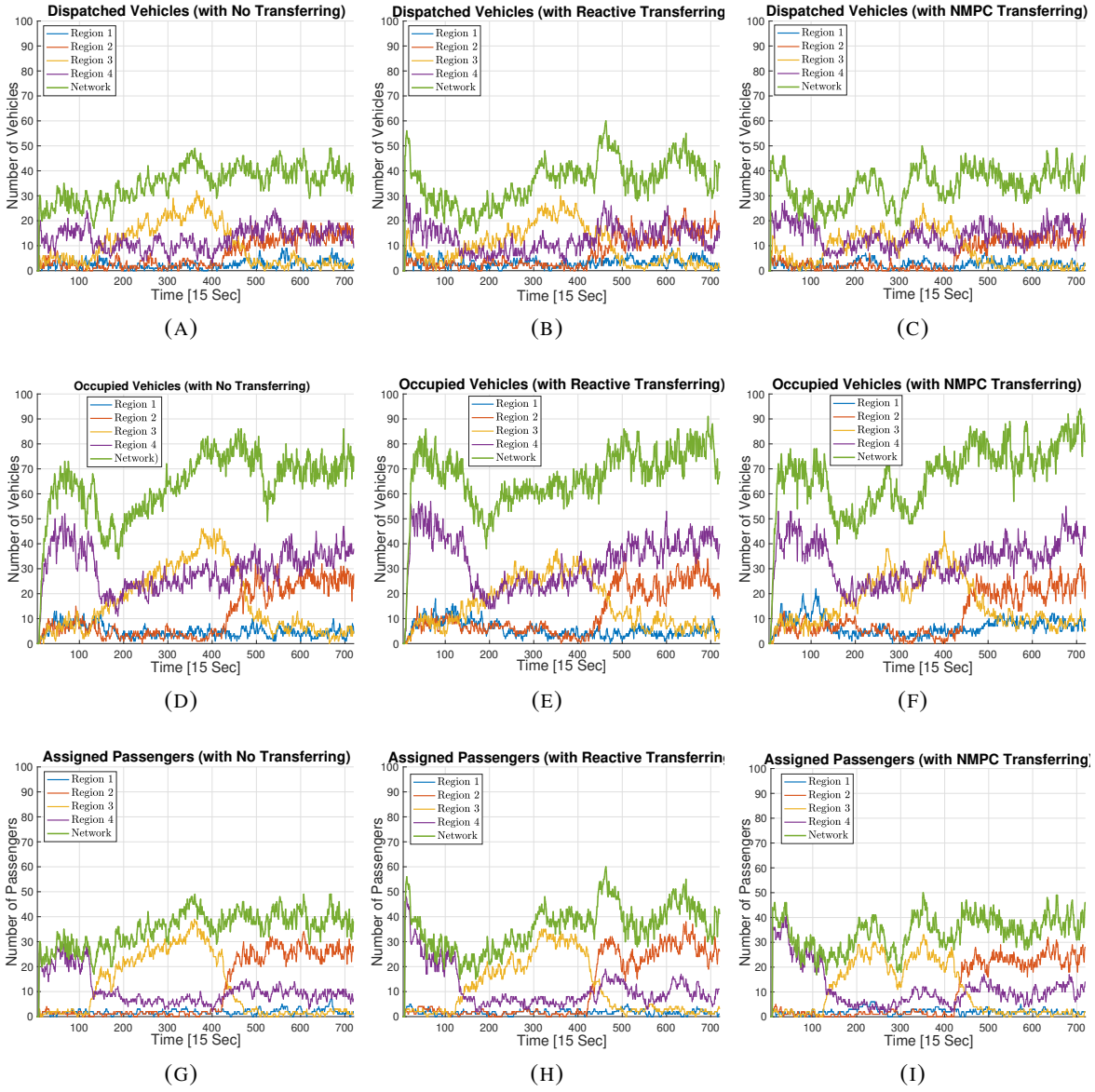
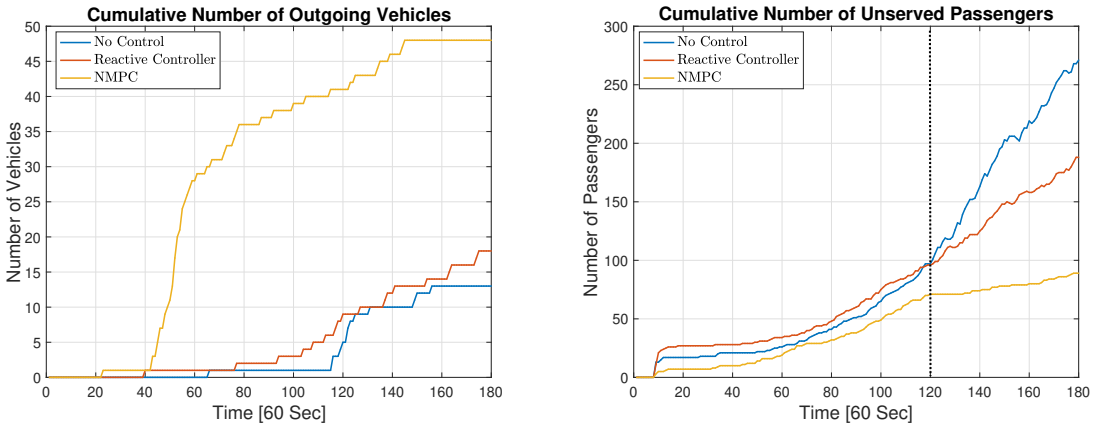


FIGURE 3.8. Figure 8: Evolution of the number of the dispatched vehicles, occupied vehicles, and assigned passengers by using different transferring methods. Dispatched vehicles (a) without transferring, (b) with reactive transferring, and (c) NMPC transferring. Occupied vehicles (d) without transferring, (e) with reactive transferring, and (f) NMPC transferring. Assigned passengers (g) without transferring, (h) with reactive transferring, and (i) NMPC transferring.

3.4.4 Served Passengers and Fleet Size Evaluation

In this part, we compare the three scenarios with respect to fleet size of the ride-sourcing and the number of the order cancellations. The stochastic impatience time threshold of each driver is randomly generated from a normal distribution with mean value of 1200 [sec] and variance of 600 [sec]. Microsimulation with no controller, reactive controller, and NMPC result in 193 [veh], 178 [veh], and 157 [veh] unique ride-sourcing vehicles required to serve the same demand of trip requests (average of five microsimulation runs). It is equal to 7.7% and 18.6% reduction in required fleet size by using the reactive controller and NMPC respectively. Figure 3.9.a represents the cumulative number of outgoing vehicles for one simulation replication (the vehicles that leave the system because of not being matched to passengers before drivers' impatience time threshold). This figure illustrates that by using NMPC, a smaller number of ride-sourcing vehicles are needed to serve the passengers. Hence, a reduction in unnecessary search trips is expected that decreases the overall network congestion level.

On the other hand, the total number of the served passengers by getting the average of five simulation replications is 1337, 1407, and 1467 by using no controller, the reactive controller, and the NMPC respectively. This shows 5.2% and 9.7% increase in the number of served passengers by using the reactive controller and NMPC respectively. Figure 3.9.b depicts the cumulative number of order cancellations. We adopt stochastic time-invariant values for passengers' impatience time threshold. In these experiments, we choose a normal distribution with a mean value of 600 [sec] and variance of 135 [sec] for the impatience time of passengers. As shown in Figure 3.9.b, the number of order cancellations by using NMPC is less than using the reactive controller or with no transfer subsystem. If we focus on the reactive controller and with no transfer scenarios, we see that after sample 120 (7200 [sec]) when the imbalance between idle vehicles and waiting passengers widens, using reactive controller reduces the number of unserved passengers. Before sample 120, by using the reactive controller, the number of unserved passengers is higher than with no transfer subsystem. However, pro-activeness of NMPC reduces the number of unserved passengers during all time instances.



(A) (B)

FIGURE 3.9. Cumulative number of (a) outgoing vehicles and (b) order cancellations because of not being matched before passengers' patience time.

Table 3.1 presents the summary of the comparison of using no controller, the reactive controller, and NMPC with respect to the number of served passengers and fleet size. The results are obtained by averaging five microsimulation replications. As illustrated in Table 3.1, the average number of served passengers by each vehicle increased 14.0% and 34.7% by using the reactive controller and the NMPC.

TABLE 3.1. The number of served passengers and operational fleet size for different scenarios of transfer subsystem. Averages of five microsimulation replications is reported. Numbers in parenthesis show percentage of the variation with respect to no transferring method.

	Average Number of Served Passengers	**Average Fleet Size**	**Average Number of Served Passengers per Vehicle**
No Transfer Subsystem	1337	193	6.93
Reactive Controller	1407 (5.2%)	178 (-7.7%)	7.90 (14.0%)
NMPC	1467 (9.7%)	157 (-18.6%)	9.34 (34.7%)

3.4.5 Delay Comparison

Table 3.2 lists the performance of different transfer subsystems with respect to the passengers' unassigned time (time that is elapsed so waiting passengers become assigned), vehicles' unassigned time (time that is elapsed so idle vehicles become dispatched), passengers' waiting

TABLE 3.2. Unassigned and waiting delays with no controller, the reactive controller, and NMPC. Results are obtained by averaging five microsimulation replications. Numbers in parenthesis show the percentage of variation with respect to the no transferring method. Lower standard deviations indicate more equitable systems.

	Passengers' Unassigned Time [sec]		**Passengers' Waiting Time [sec]**		**Vehicles' Unassigned Time [sec]**		**Vehicles' Waiting Time [sec]**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
No Transfer Subsystem	36	57	166	124	161	337	749	1200
Reactive Controller	34 (-5.5%)	56	151.2 (-8.9%)	122	228 (40.7%)	501	736 (-1.9%)	1102
NMPC	28 (-20.4%)	47	146 (-12.4%)	116	174 (7.8%)	316	685 (-8.8%)	970

time (time that is elapsed so waiting passengers become on-board), and vehicles' waiting time (time that is elapsed so idle vehicles become occupied) for successful pick ups. By using the reactive controller and NMPC, the vehicles' unassigned time increases because transferring trips postpone immediate matching of the (excess) number of the idle vehicles in favor of proactive balancing between new arriving waiting passengers and idle vehicles. Noteworthy, the passengers' and vehicles' waiting time decrease. Because, instead of instantaneous matching, the transfer subsystem recommends vehicles to reposition to other regions that are suitable for future matching (i.e. vehicles being closer to new arriving waiting passengers). NMPC results in lower standard deviations of the reported measures that indicates the higher equity of the ride-sourcing system among drivers and among passengers. NMPC results in lower standard deviations of the reported measures that indicates the higher equity of the ride-sourcing system among drivers and among passengers.

Ultimately, the NMPC has better performance in terms of reducing the number of the waiting passengers and increasing the rate of occupied vehicles in comparison with the reactive controller because the NMPC (i) considers current location and *future destination* of different modes of vehicles and passengers and (ii) incorporates the effect of transferring for reducing the *predicted* number of the waiting passengers and transferred vehicles.

3.4.6 Drivers Incentives

In this part, we focus on the effect of the reactive controller and NMPC on travelled distance of the transferred and dispatched vehicles. We show that drivers (contractors) by following

the transferring commands of the ride-sourcing system reduce their Guided Vacant Cruising Distance (GVCD), that is the sum of the transferred and dispatched cruising distances. Figures 3.10.b and 3.10.c depict the normalised histogram of the GVCD of the trips that the vehicle has been transferred prior to being matched. Figure 3.10.a illustrates the histogram of the GVCD with no transferring. In this figure, GVCD is reported for trips of the vehicles that has received at least one transferring command in NMPC scenario. (This is to compare the vacant trips of transferred vehicles in days when the transferring subsystem is active versus days transferring is not active.)

Figures 3.10.b and 3.10.c depict the mean values of the GVCD for the reactive controller and NMPC are 534 [m] and 475 [m]. Hence, pro-activeness of NMPC reduces the GVCD of the trips of the transferred vehicles by 12.4%. Moreover, by using the reactive controller and the NMPC, 95% of the trips has GVCD less than 1094 [m] and 920 [m] respectively. This is equal to 13% improvements in reducing the matching distances without considering the effects of outliers. With no transferring scenario, Figure 3.10.a, the standard deviation of the GVCD of the trips increases by 62.2%, which demonstrates existence of the wild goose chase problem. In addition, the mean value of the GVCD with no transferring increases by 17.4% in comparison with the NMPC.

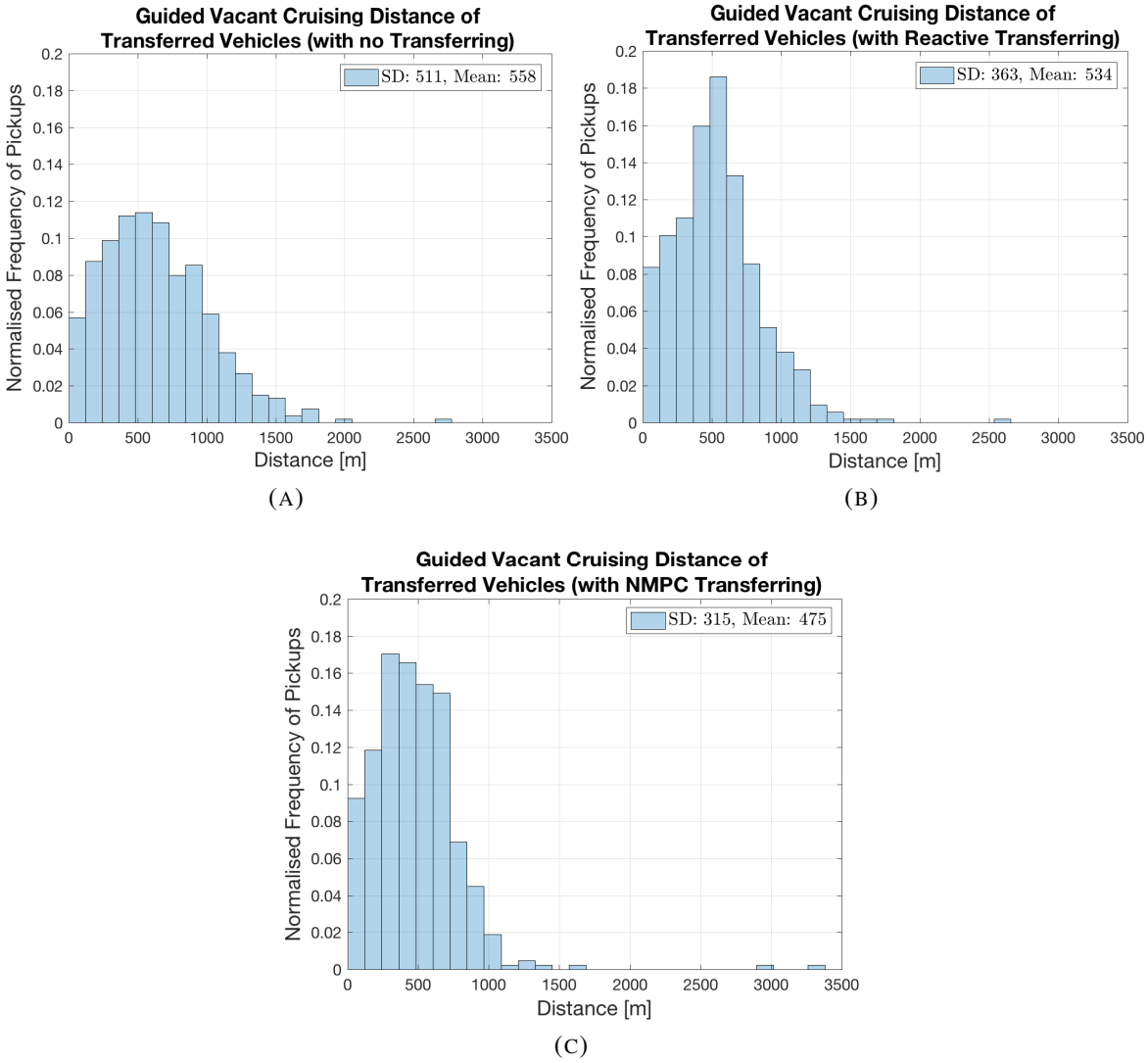


FIGURE 3.10. Normalised histogram of GVCD of trips of (a) the vehicles that at least have received a transferring command by NMPC controller; however, they operate in no control scenario, (b) the vehicles that at least receive a transferring command by the reactive controller, and (c) the vehicles that at least receive a transferring command by the NMPC.

Figure 3.11 illustrates the normalised histogram of the GVCD of all the trips whether the vehicle receives transferring command or not. We can see the distribution of the GVCD is skewed to right by using NMPC as the transferring method in comparison with no controller and the reactive controller. The mean value of the GVCD is 573 [m], 508 [m], and 491 [m] by using no controller, the reactive controller, and the NMPC respectively. It is equal to 14.3%

and 3.35% reduction of GVCD by using NMPC in comparison with no controller and the reactive controller. These results demonstrate that the drivers by following the transferring recommendations reduce their total GVCD which is their personal (selfish) gain. Accordingly, it can be expected that rational drivers would follow the repositioning recommendations over time.

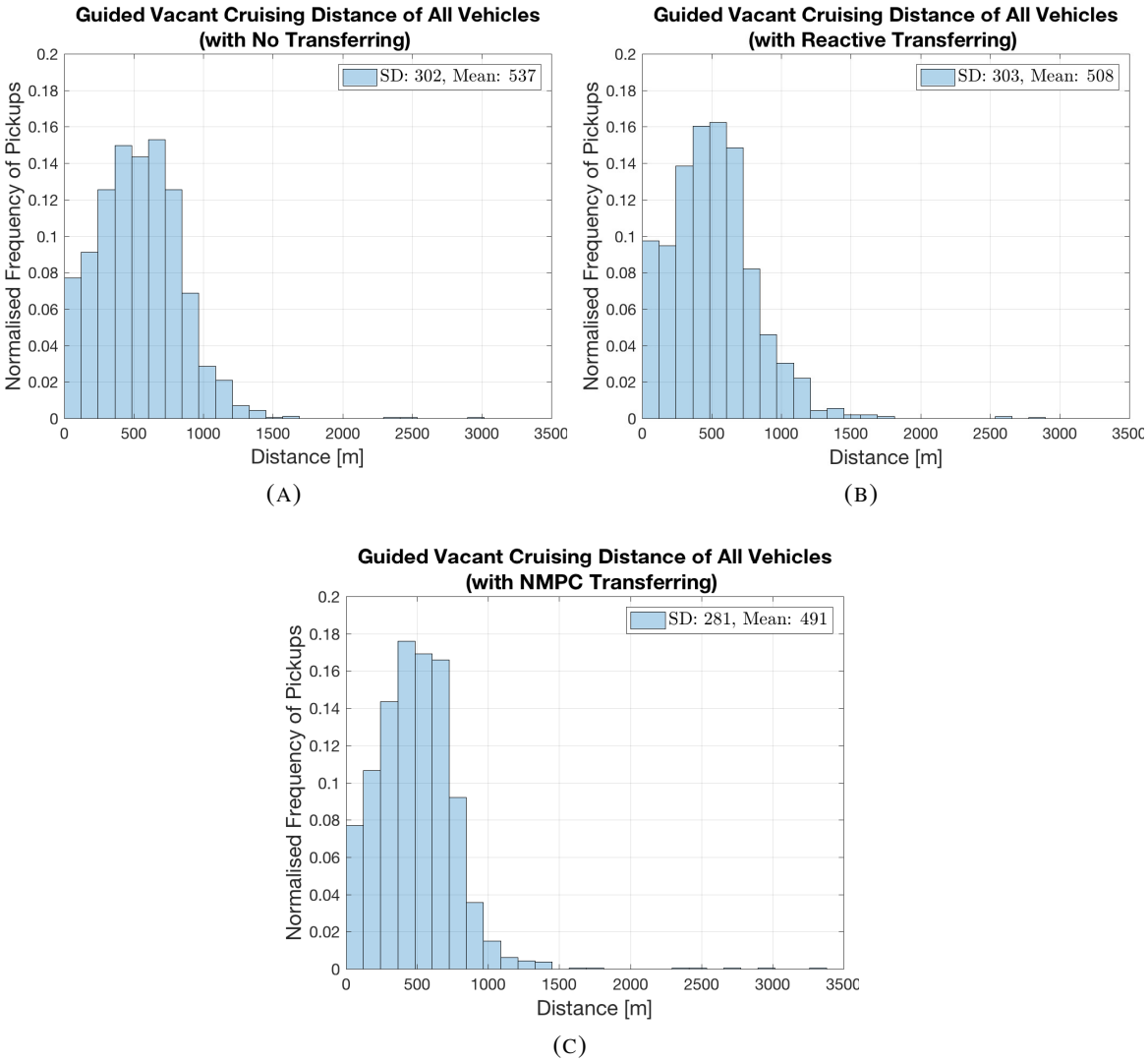


FIGURE 3.11. Normalised histogram of GVCD of trips of all the vehicles whether they receive the transferring command or not with (a) no controller, (b) the reactive controller, and (c) the NMPC.

3.5 Summary and Future Work

This article has proposed a model-based nonlinear model predictive controller (NMPC) to proactively reposition the idle vehicles in the network to balance the demand of waiting passengers and the supply of vehicles in ride-sourcing systems. The proposed NMPC is designed based on the model in Chapter 2 to reposition the vehicles to the regions with a higher probability of being matched to new passengers. To increase the controllability of the model and pro-activeness of the transferring, closed-form formulations are suggested to estimate the rates of vehicle-passenger matchings during the prediction horizons. The NMPC utilises the estimated values of the vehicle-passenger matchings and the model in a state-space form to predict the evolution of the ride-sourcing vehicles (i.e. idle, transferred, dispatched, and occupied vehicles) and passengers (i.e. waiting and assigned passengers). The NMPC determines the optimum transferring rates by solving a constrained finite-horizon optimal control problem to minimize the total unassigned time of the vehicles and passengers.

The effectiveness of the NMPC as the transferring method has been compared with no controller and a reactive controller scenarios using traffic microsimulation experiments. We have demonstrated that utilizing the proposed NMPC for repositioning the idle vehicles improves equality of the ride-sourcing system for both passengers and drivers. We showed that the NMPC has (i) increased the total number of the served passengers by a smaller fleet size, (ii) reduced the passengers' unassigned and waiting times, and (iii) reduced the vehicles' waiting times. In addition to the network-level improvements, we have shown that using NMPC incentivises the drivers by reducing their guided vacant cruising distance.

A future research direction is to consider the effect of transfer frequencies on (i) transferring rates, (ii) matching distances, and (iii) matching frequencies. Incorporating the transferring frequencies in modeling is a possible solution that results in a hybrid or a switching system in which the transfer frequencies can be obtained dynamically as a function of the evolution of the vehicles and passengers. Another challenging research is extending the proposed matching and transferring methods for ride-splitting and ride-sharing systems. In addition, transferring of idle vehicles can be considered integrated with surge pricing methods to

curb the supply-demand mismatch. In this Chapter, arrival rate of waiting passengers is a time-varying exogenous factor. Considering elasticity of this factor to repositioning, total delay, average passengers' unassigned time or other long-term variables is a research priority. In addition, taking into account the supply elasticity can be utilised for investigating detailed monetary incentivization of individual drivers and their compliance rate in following platform transferring recommendations. Furthermore, studying the effect of competition between multi-homing contractors on supply of vehicles is another future research direction. Integration of ride-sourcing fleet information as probe vehicles data to estimate region-level traffic speed based on MFD and incorporating such traffic dynamics into the model and controller is also a research direction for further studies.

Appendix: Vehicle Repositioning and Matching without Discarding

To isolate the effect of the proposed proactive transferring method from the proposed adaptive spatio-temporal matching in Chapter 2, we employ the optimal myopic method with fixed matching interval of 30 [sec] and no discarding. This would let scrutinize the transferring method in isolation. Note that any transferring methods requires a pool of idle vehicles that can be potentially transferred to balance the supply and demand in different parts of the network. Thus, a matching method such as optimal myopic that exhausts all the supply resources of idle vehicles for shorter passengers' matching times, in expense of wild goose chase of long-distance matchings, would do a disservice to any transferring method (i.e. reducing the number of the idle vehicles reduces the effectiveness of the transferring method).

Table 3.3 shows the effect of the transferring subsystem on average number of served passengers and average fleet size (i.e., ride-sourcing vehicles). Table 3.4 lists the performance of the 3 transferring scenarios on vehicles' and passengers' delay. The numbers in these tables are obtained by using optimal myopic matching method with fixed matching interval of 30 [sec] and no discarding. As shown in Table 3.3, repositioning mildly increases the average of number of served passengers and reduces the fleet size that is required to serve the

passengers. The reason is that the optimal myopic matching method consumes all the supply resources (i.e., idle vehicles). By comparing Table 3.4 and Table 2.1, using optimal myopic with no discarding matching method instead of the proposed adaptive spatio-temporal method increases the passengers' and vehicles' unassigned times and waiting times. Furthermore, these results demonstrate that optimal myopic matching method with no discarding reduces the performance of the transferring subsystem.

TABLE 3.3. The number of served passengers and operational fleet size for different scenarios of transfer subsystem and using optimal myopic matching method with fixed matching interval of 30 [sec] and no discarding. Averages of five microsimulation replications is reported. Numbers in parenthesis show percentage of the variation with respect to no transferring method.

	Average Number of Served Passengers	**Average Fleet Size**	**Average Number of Served Passengers per Vehicle**
No Transfer Subsystem	1097	218	5.03
Reactive Controller	1101 (0.4%)	212 (-1.8%)	5.19 (3.2%)
NMPC	1105 (0.7%)	212 (-1.8%)	5.21 (3.6%)

TABLE 3.4. Unassigned and waiting delays with no controller, the reactive controller, and NMPC using optimal myopic matching method with fixed matching interval of 30 [sec] and no discarding. Results are obtained by averaging five microsimulation replications. Numbers in parenthesis show the percentage of variation with respect to the no transferring method.

	Passengers' Unassigned Time [sec]		**Passengers' Waiting Time [sec]**		**Vehicles' Unassigned Time [sec]**		**Vehicles' Waiting Time [sec]**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
No Transfer Subsystem	40.4	54	197.0	137	148.0	329	793	1021
Reactive Controller	39.5 (-2.2%)	51	192.2 (-2.4%)	134	153.0 (3.3%)	339	779 (-1.7%)	1009
NMPC	39.1 (-3.2%)	51	189.9 (-3.6%)	131	155.1 (4.8%)	337	777 (-2.0%)	1005

CHAPTER 4

Conclusion and Future Work

In this thesis, we proposed a holistic framework that contains three components: vehicle-passenger matching method, modelling for predicting the evolution of the ride-sourcing system's state (i.e. idle vehicles, transferred vehicles, dispatched vehicles, occupied vehicles, waiting passengers, assigned passengers, and on-board passengers) in each region, and repositioning idle vehicles to the locations with higher probabilities of being matched to the waiting passengers. Besides, the detailed interactions between the aforementioned subsystems have been elaborated. The ride-sourcing companies can utilize the holistic ride-sourcing solution or each individual component to improve the level of service of their system. Besides, the proposed model can be used by transport decision makers to predict the effect of ride-sourcing companies matching/transferring policies on traffic congestion.

In the proposed framework, three clocks have been run in parallel for executing vehicle-passengers matching method, triggering transferring subsystem, and prediction of the state of ride-sourcing system. At each matching frequency, a subset of the idle and transferred vehicles that are selected by the adaptive spatio-temporal matching method is dispatched to the waiting passengers. At each transferring time, the remaining idle vehicles are relocated to the locations with higher probabilities of being dispatched to the waiting passengers to minimize the total waiting time of the vehicles and passengers up to the defined prediction horizon that is discretized with respect to prediction sample time.

The proposed model was developed based on macroscopic fundamental diagrams and Cobb-Douglas meeting functions to simulate the ride-sourcing system behavior without getting feedback for correcting the model mismatch. The model has been introduced to predict the evolution of the different status of vehicles and passengers in each region of the network. The

non-equilibrium based model is valid under high variation of exogenous inputs (i.e. rate of new idle vehicles and the arrival rate of waiting passengers) and it overcomes the limitations of steady-state modeling. The proposed model considers manipulating variables (i.e. control inputs) for repositioning idle vehicles, impatient passengers and drivers, and rate of the vehicle-passenger matching method. The model can be utilised as a stand-alone subsystem for the initial evaluation of the vehicle-passenger matching method or repositioning method without using microsimulation. The validity of the macroscopic model has been evaluated with microsimulation experiments.

Another subsystem of the holistic ride-sourcing framework is the vehicle-passenger matching method. We proposed an adaptive spatio-temporal matching method to dynamically determines optimum next matching interval and maximum matching distance with respect to minimizing expected passengers' waiting time. The joint effect of the matching intervals and the maximum matching distance and the level of congestion is reflected in the proposed method. The microsimulation experiments reveal that the adaptive spatio-temporal matching method reduces vehicles' and passengers' delay in comparison with the perfect and greedy matching methods.

The third component of the holistic ride-sourcing framework is transferring subsystem. A nonlinear model predictive controller (NMPC) was proposed for reposition idle vehicles to balance the supply of vehicles and the demand of the waiting passengers in each region of the network. The NMPC that is designed based on the model relocates idle vehicles proactively to minimize the cost function that is defined based on a definite integral of waiting time of vehicles and passengers with an upper bound equal to the defined prediction horizon. Microsimulation reveals that the NMPC has increased the total number of the served passengers with a smaller fleet size, reduced the passengers' unassigned and waiting times, and reduced the vehicles' waiting times. Furthermore, we have shown utilizing the NMPC incentives for majority of vehicles' drivers by reducing guided vacant cruising distance of the vehicles that at least received a transferring command.

This thesis has proposed a model-based nonlinear model predictive controller (NMPC) to proactively reposition the idle vehicles in the network to balance the demand of waiting

passengers and the supply of vehicles in ride-sourcing systems. The proposed NMPC is designed based on the model in Part I to reposition the vehicles to the regions with a higher probability of being matched to new passengers. To increase the controllability of the model and pro-activeness of the transferring, closed-form formulations are suggested to estimate the rates of vehicle-passenger matchings during the prediction horizons. The NMPC utilises the estimated values of the vehicle-passenger matchings and the model in a state-space form to predict the evolution of the ride-sourcing vehicles (i.e. idle, transferred, dispatched, and occupied vehicles) and passengers (i.e. waiting and assigned passengers). The NMPC determines the optimum transferring rates by solving a constrained finite-horizon optimal control problem to minimize the total unassigned time of the vehicles and passengers.

Bibliography

Agatz, Niels et al. (2011). 'Dynamic ride-sharing: A simulation study in metro Atlanta'. In: *Procedia-Social and Behavioral Sciences* 17, pp. 532–550.

Agatz, Niels et al. (2012). 'Optimization for dynamic ride-sharing: A review'. In: *European Journal of Operational Research* 223.2, pp. 295–303.

Alisoltani, Negin, Ludovic Leclercq and Mahdi Zargayouna (2021). 'Can dynamic ride-sharing reduce traffic congestion?' In: *Transportation research part B: methodological* 145, pp. 212–246.

Alonso-Mora, Javier et al. (2017). 'On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment'. In: *Proceedings of the National Academy of Sciences* 114.3, pp. 462–467.

Amar, Haitham M and Otman A Basir (2018). 'A Game Theoretic Solution for the Territory Sharing Problem in Social Taxi Networks'. In: *IEEE Transactions on Intelligent Transportation Systems*.

Averch, Harvey and Leland L Johnson (1962). 'Behavior of the firm under regulatory constraint'. In: *The American Economic Review* 52.5, pp. 1052–1069.

Barketau, Maksim, Erwin Pesch and Yakov Shafransky (2015). 'Minimizing maximum weight of subsets of a maximum matching in a bipartite graph'. In: *Discrete Applied Mathematics* 196, pp. 4–19.

Battifarano, Matthew and Zhen Sean Qian (2019). 'Predicting real-time surge pricing of ride-sourcing companies'. In: *Transportation Research Part C: Emerging Technologies* 107, pp. 444–462.

Beesley, Michael E (1973). 'Regulation of taxis'. In: *The economic journal* 83.329, pp. 150–172.

Bertsimas, Dimitris, Patrick Jaillet and Sébastien Martin (2019). 'Online vehicle routing: The edge of optimization in large-scale applications'. In: *Operations Research* 67.1, pp. 143–162.

Braverman, Anton et al. (2019). 'Empty-car routing in ridesharing systems'. In: *Operations Research* 67.5, pp. 1437–1452.

Byrd, Richard H, Jorge Nocedal and Richard A Waltz (2006). 'K nitro: An integrated package for nonlinear optimization'. In: *Large-scale nonlinear optimization*. Springer, pp. 35–59.

Castillo, Juan Camilo, Dan Knoepfle and Glen Weyl (2017). 'Surge pricing solves the wild goose chase'. In: *Proceedings of the 2017 ACM Conference on Economics and Computation*, pp. 241–242.

Chamberlin, Edward Hastings (1949). *Theory of monopolistic competition: A re-orientation of the theory of value*. Oxford University Press, London.

Coffman, Richard B and Chanoch Shreiber (1977). 'The Economic Reasons for Price and Entry Regulation of Taxicabs (Comment and Rejoinder)'. In: *Journal of Transport Economics and Policy*, pp. 288–304.

Dandl, Florian et al. (2019). 'Evaluating the impact of spatio-temporal demand forecast aggregation on the operational performance of shared autonomous mobility fleets'. In: *Transportation* 46.6, pp. 1975–1996.

De Vany, Arthur S (1975). 'Capacity utilization under alternative regulatory restraints: an analysis of taxi markets'. In: *Journal of political economy* 83.1, pp. 83–94.

Douglas, George W (1972a). 'Price regulation and optimal service standards: The taxicab industry'. In: *Journal of Transport Economics and Policy*, pp. 116–127.

— (1972b). 'Price regulation and optimal service standards: The taxicab industry'. In: *Journal of Transport Economics and Policy*, pp. 116–127.

Friedman, Milton (1962). *Price theory, a provisional text.* Tech. rep.

Gan, Jiarui et al. (2013). 'Optimal Pricing for Improving Efficiency of Taxi Systems.' In: *IJCAI*, pp. 2811–2818.

Geroliminis, Nikolas and Carlos F Daganzo (2008). 'Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings'. In: *Transportation Research Part B: Methodological* 42.9, pp. 759–770.

Haddad, Jack and Zhengfei Zheng (2020). 'Adaptive perimeter control for multi-region accumulation-based models with state delays'. In: *Transportation Research Part B: Methodological* 137, pp. 133–153.

Ho, Sin C et al. (2018). 'A survey of dial-a-ride problems: Literature review and recent developments'. In: *Transportation Research Part B: Methodological* 111, pp. 395–421.

Hörl, Sebastian et al. (2019). 'Fleet operational policies for automated mobility: A simulation assessment for Zurich'. In: *Transportation Research Part C: Emerging Technologies* 102, pp. 20–31.

Hu, Liang and Jing Dong (2020). 'An Artificial-Neural-Network-Based Model for Real-Time Dispatching of Electric Autonomous Taxis'. In: *IEEE Transactions on Intelligent Transportation Systems*.

Huang, Zhenhua et al. (2017). 'PRACE: A Taxi Recommender for Finding Passengers with Deep Learning Approaches'. In: *International Conference on Intelligent Computing*. Springer, pp. 759–770.

Illgen, Stefan and Michael Höck (2019). 'Literature review of the vehicle relocation problem in one-way car sharing networks'. In: *Transportation Research Part B: Methodological* 120, pp. 193–204.

Ingole, Deepak, Guilhem Mariotte and Ludovic Leclercq (2020). 'Perimeter gating control and citywide dynamic user equilibrium: A macroscopic modeling framework'. In: *Transportation research part C: emerging technologies* 111, pp. 22–49.

Ke, Jintao et al. (2017). 'Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach'. In: *Transportation Research Part C: Emerging Technologies* 85, pp. 591–608.

Ke, Jintao et al. (2018). 'Hexagon-based convolutional neural network for supply-demand forecasting of ride-sourcing services'. In: *IEEE Transactions on Intelligent Transportation Systems* 20.11, pp. 4160–4173.

Kouvelas, Anastasios, Mohammadreza Saeedmanesh and Nikolas Geroliminis (2017). 'Enhancing model-based feedback perimeter control with data-driven online adaptive optimization'. In: *Transportation Research Part B: Methodological* 96, pp. 26–45.

Li, Ye, Reza Mohajerpoor and Mohsen Ramezani (2021). 'Perimeter control with real-time location-varying cordon'. In: *Transportation Research Part B: Methodological* 150, pp. 101–120.

Lu, Yu et al. (2018). 'An intelligent system for taxi service: Analysis, prediction and visualization'. In: *AI Communications* Preprint, pp. 1–14.

Maciejewski, Michal, Joschka Bischoff and Kai Nagel (2016). 'An assignment-based approach to efficient real-time city-scale taxi dispatching'. In: *IEEE Intelligent Systems* 31.1, pp. 68–77.

Mahmassani, Hani S, Meead Saberi and Ali Zockaie (2013). 'Urban network gridlock: Theory, characteristics, and dynamics'. In: *Procedia-Social and Behavioral Sciences* 80, pp. 79–98.

Manski, Charles F and J David Wright (1967). *Nature of equilibrium in the market for taxi services*. Tech. rep.

Narayanan, Santhanakrishnan, Emmanouil Chaniotakis and Constantinos Antoniou (2020). 'Shared autonomous vehicle services: A comprehensive review'. In: *Transportation Research Part C: Emerging Technologies* 111, pp. 255–293.

Nourinejad, Mehdi and Mohsen Ramezani (2019). 'Ride-Sourcing modeling and pricing in non-equilibrium two-sided markets'. In: *Transportation Research Part B: Methodological*.

Nourinejad, Mehdi et al. (2015). 'Vehicle relocation and staff rebalancing in one-way carsharing systems'. In: *Transportation Research Part E: Logistics and Transportation Review* 81, pp. 98–113.

Orr, Daniel (1969a). 'The "Taxicab problem": A proposed solution'. In: *Journal of Political Economy* 77.1, pp. 141–147.

— (1969b). 'The" Taxicab problem": A proposed solution'. In: *Journal of Political Economy* 77.1, pp. 141–147.

Pavone, Marco et al. (2012). 'Robotic load balancing for mobility-on-demand systems'. In: *The International Journal of Robotics Research* 31.7, pp. 839–854.

Qian, Xinwu and Satish V Ukkusuri (2017). 'Taxi market equilibrium with third-party hailing service'. In: *Transportation Research Part B: Methodological* 100, pp. 43–63.

Ramezani, Mohsen, Jack Haddad and Nikolas Geroliminis (2015). 'Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control'. In: *Transportation Research Part B: Methodological* 74, pp. 1–19.

Ramezani, Mohsen and Mehdi Nourinejad (2018a). 'Dynamic modeling and control of taxi services in large-scale urban networks: A macroscopic approach'. In: *Transportation Research Part C: Emerging Technologies* 94, pp. 203–219.

— (2018b). 'Dynamic modeling and control of taxi services in large-scale urban networks: A macroscopic approach'. In: *Transportation Research Part C: Emerging Technologies* 94, pp. 203–219.

Saeedmanesh, Mohammadreza and Nikolas Geroliminis (2016). 'Clustering of heterogeneous networks with directional flows based on "Snake" similarities'. In: *Transportation Research Part B: Methodological* 91, pp. 250–269.

Salanova, Josep Maria et al. (2011). 'A review of the modeling of taxi services'. In: *Procedia-Social and Behavioral Sciences* 20, pp. 150–161.

Salhab, Rabih, Jerome Le Ny and Roland P Malhamé (2017). 'A dynamic ride-sourcing game with many drivers'. In: *Communication, Control, and Computing (Allerton), 2017 55th Annual Allerton Conference on*. IEEE, pp. 770–775.

Santi, Paolo et al. (2014). 'Quantifying the benefits of vehicle pooling with shareability networks'. In: *Proceedings of the National Academy of Sciences* 111.37, pp. 13290–13294.

Sayarshad, Hamid R and Joseph YJ Chow (2017). 'Non-myopic relocation of idle mobility-on-demand vehicles as a dynamic location-allocation-queueing problem'. In: *Transportation Research Part E: Logistics and Transportation Review* 106, pp. 60–77.

Shreiber, Chanoch (1975). 'The Economic Reasons for Price and Entry Regugation of Taxicabs'. In: *Journal of Transport Economics and Policy*, pp. 268–279.

Vazifeh, Mohammed M et al. (2018). 'Addressing the minimum fleet problem in on-demand urban mobility'. In: *Nature* 557.7706, p. 534.

Vazirani, Vijay V (1994). 'A theory of alternating paths and blossoms for proving correctness of the $O(\sqrt{V}E)$ general graph maximum matching algorithm'. In: *Combinatorica* 14.1, pp. 71–109.

Wächter, Andreas and Lorenz T Biegler (2006). 'On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming'. In: *Mathematical programming* 106.1, pp. 25–57.

Waltz, Richard A et al. (2006). 'An interior algorithm for nonlinear optimization that combines line search and trust region steps'. In: *Mathematical programming* 107.3, pp. 391–408.

Wang, Dong et al. (2017a). 'DeepSD: supply-demand prediction for online car-hailing services using deep neural networks'. In: *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*. IEEE, pp. 243–254.

Wang, Hai and Hai Yang (2019). 'Ridesourcing systems: A framework and review'. In: *Transportation Research Part B: Methodological* 129, pp. 122–155.

Wang, Xing, Niels Agatz and Alan Erera (2017b). 'Stable matching for dynamic ride-sharing systems'. In: *Transportation Science* 52.4, pp. 850–867.

Williams, David J (1980). 'The economic reasons for price and entry regulation of taxicabs: A comment'. In: *Journal of Transport Economics and Policy* 14.1, pp. 105–112.

Wong, KI, Sze Chun Wong and Hai Yang (2001). 'Modeling urban taxi services in congested road networks with elastic demand'. In: *Transportation Research Part B: Methodological* 35.9, pp. 819–842.

Wong, KI et al. (2002). 'A sensitivity-based solution algorithm for the network model of urban taxi services'. In: *Transportation and Traffic Theory in the 21st Century: Proceedings of the 15th International Symposium on Transportation and Traffic Theory, Adelaide, Australia, 16-18 July 2002*. Emerald Group Publishing Limited, pp. 23–42.

Wong, KI et al. (2005). 'Modeling the bilateral micro-searching behavior for urban taxi services using the absorbing markov chain approach'. In: *Journal of advanced transportation* 39.1, pp. 81–104.

Wong, RCP, WY Szeto and SC Wong (2014). 'A cell-based logit-opportunity taxi customer-search model'. In: *Transportation Research Part C: Emerging Technologies* 48, pp. 84–96.

— (2015). 'A two-stage approach to modeling vacant taxi movements'. In: *Transportation Research Procedia* 7, pp. 254–275.

Wu, Xin et al. (2018). ‘Hierarchical travel demand estimation using multiple data sources: A forward and backward propagation algorithmic framework on a layered computational graph’. In: *Transportation Research Part C: Emerging Technologies* 96, pp. 321–346.

Xu, Jianmin et al. (1999). ‘Modeling level of urban taxi services using neural network’. In: *Journal of Transportation Engineering* 125.3, pp. 216–223.

Xu, Zhengtian, Yafeng Yin and Jieping Ye (2020). ‘On the supply curve of ride-hailing systems’. In: *Transportation Research Part B: Methodological* 132, pp. 29–43.

Yang, Hai and SC Wong (1998). ‘A network model of urban taxi services’. In: *Transportation Research Part B: Methodological* 32.4, pp. 235–246.

Yang, Hai, Sze Chun Wong and Kelvin I Wong (2002a). ‘Demand–supply equilibrium of taxi services in a network under competition and regulation’. In: *Transportation Research Part B: Methodological* 36.9, pp. 799–819.

Yang, Hai and Teng Yang (2011). ‘Equilibrium properties of taxi markets with search frictions’. In: *Transportation Research Part B: Methodological* 45.4, pp. 696–713.

Yang, Hai et al. (2000). ‘A macroscopic taxi model for passenger demand, taxi utilization and level of services’. In: *Transportation* 27.3, pp. 317–340.

Yang, Hai et al. (2002b). ‘Modeling Urben Taxi Services: A Literature Survey and An Analytical Example’. In: *Advanced Modeling for Transit Operations and Service Planning*.

Yang, Hai et al. (2010a). ‘Equilibria of bilateral taxi–customer searching and meeting on networks’. In: *Transportation Research Part B: Methodological* 44.8-9, pp. 1067–1083.

Yang, Hai et al. (2010b). ‘Nonlinear pricing of taxi services’. In: *Transportation Research Part A: Policy and Practice* 44.5, pp. 337–348.

Yang, Hai et al. (2020). ‘Optimizing matching time interval and matching radius in on-demand ride-sourcing markets’. In: *Transportation Research Part B: Methodological* 131, pp. 84–105.

Yildirimoglu, Mehmet, Isik Ilber Sirmatel and Nikolas Geroliminis (2018). ‘Hierarchical control of heterogeneous large-scale urban road networks via path assignment and regional route guidance’. In: *Transportation Research Part B: Methodological* 118, pp. 106–123.

Zha, Liteng, Yafeng Yin and Yuchuan Du (2018a). 'Surge pricing and labor supply in the ride-sourcing market'. In: *Transportation Research Part B: Methodological* 117, pp. 708–722.

Zha, Liteng, Yafeng Yin and Zhengtian Xu (2018b). 'Geometric matching and spatial pricing in ride-sourcing markets'. In: *Transportation Research Part C: Emerging Technologies* 92, pp. 58–75.

Zha, Liteng, Yafeng Yin and Hai Yang (2016). 'Economic analysis of ride-sourcing markets'. In: *Transportation Research Part C: Emerging Technologies* 71, pp. 249–266.

Zhan, Xianyuan, Xinwu Qian and Satish V Ukkusuri (2016). 'A graph-based approach to measuring the efficiency of an urban taxi service system'. In: *IEEE Transactions on Intelligent Transportation Systems* 17.9, pp. 2479–2489.

Zhang, Kenan et al. (2019). 'An efficiency paradox of uberization'. In: *Available at SSRN 3462912*.

Zhang, Rick and Marco Pavone (2016). 'Control of robotic mobility-on-demand systems: a queueing-theoretical perspective'. In: *The International Journal of Robotics Research* 35.1-3, pp. 186–203.

Zheng, Nan et al. (2012). 'A dynamic cordon pricing scheme combining the macroscopic fundamental diagram and an agent-based traffic model'. In: *Transportation Research Part A: Policy and Practice* 46.8, pp. 1291–1303.

Zuniga-Garcia, Natalia et al. (2020). 'Evaluation of ride-sourcing search frictions and driver productivity: A spatial denoising approach'. In: *Transportation Research Part C: Emerging Technologies* 110, pp. 346–367.