BIKE SHARE SYSTEM MODELING: TRAVEL PATTERNS, ENVIRONMENTAL BENEFITS, SYSTEM EXPANSION AND IMPACTS OF SYSTEM TYPES

by

Zhaoyu Kou

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Industrial Engineering

West Lafayette, Indiana

December 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL

STATEMENT OF COMMITTEE APPROVAL

Dr. Hua Cai, Chair

School of Industrial Engineering

Environmental and Ecological Engineering

Dr. Gemma Berenguer

Department of Business Administration

Universidad Carlos III de Madrid, Spain

Dr. Seokcheon Lee

School of Industrial Engineering

Dr. Shimon Y. Nof

School of Industrial Engineering

Approved by:

Dr. Abhijit Deshmukh

This dissertation is dedicated to my loved parents and wife
who supported me in all things great and small

ACKNOWLEDGMENTS

First, I would like to express sincere gratitude to my advisor, Dr. Hua Cai for her guidance, supports, and encouragement during my PhD studies. With her instructions, I was able to explore different research problems, broad my knowledge, and improve my critical and creative thinking abilities.

Second, I would like to thank all committee members: Dr. Gemma Berenguer, Dr. Seokcheon Lee, and Dr. Shimon Y.Nof. I appreciate their time in reviewing my work and their insightful comments and suggestions that helped improve my work.

Third, I would like to thank the financial support from Hugh W. and Edna M. Donnan Fellowship (2019-2020) from the School of Industrial Engineering and the summer research grant (2017) from Environmental and Ecological Engineering at Purdue University.

Last, I would like to thank all the members of the Urban Sustainability Modeling & Analysis Research Team (uSMART) for their suggestions and supports.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BSS	Bike Share System
BS-EREM	Bike Share Emission Reduction Estimation Model
CCDF	Complementary Cumulative Distribution Function
CHTS	California Household Travel Survey
CO_2-eq	Carbon Dioxide Equivalent
DENS	Station densities in concentric distance bands at 0.1-mile interval
DENS2	Station densities in concentric distance bands with two bands: 0-0.5 miles and 0.5-3.1 miles
ECC	Spatial eccentricity of the stations
ECCQ	Spatial eccentricity quantile of the stations
ECDF	Empirical Cumulative Distribution Function
EXT	External data including socio-demographic and point-of-interest (POI) data
GEO	Latitude and Longitude of the stations
GHG	Greenhouse gas
MAPE	Mean absolute percentage error
NHTS	National Household Travel Survey
PDF	Probability Density Function
POI	Point-of-interest
RMSE	Root-mean-square error
SEQEM	Spatial Eccentricity Quantile based Ensemble Model
TSP	Traveling Salesman Problem

ABSTRACT

As an emerging shared mobility option, bike share has the potential to improve transportation sustainability. Understanding the mobility patterns, environmental benefits, and impact of system changes helps improve bike share systems (BSSs). However, existing literature has the following gaps: (1) there is a lack in detailed understanding about bike share's travel patterns; (2) few studies have considered the heterogeneous travel mode choices to quantify BSSs' environmental impacts; (3) the station interactions in the system expansion process are not well studied; and (4) there is a critical need for models that can quantify system performances of different types of BSSs from the perspectives of both users and operators.

This dissertation aims to address these gaps to provide better understanding of the travel patterns, benefits, and impacts of system changes in BSSs to assist the policy making and development of BSSs. To achieve this objective, various modeling frameworks and methods were developed. (1) The statistical property of bike share trip distance and duration are first analyzed to provide fundamental basis for the modelling of BSSs. (2) A Bike Share Emission Reduction Estimation Model (BS-EREM) is proposed to quantify the environmental benefits from BSSs. The BS-EREM estimates the transportation modes substituted by bike share trips, with the consideration of heterogeneous travel mode choices. The GHG emission reductions contributed by BSSs in eight case study cities were then evaluated using BS-EREM. (3) For system expansion, the competition/complement interactions between stations are revealed using a piecewise regression model. The study also shows that incorporating features about such interactions significantly improves the demand predictions for system expansion. A Spatial Eccentricity Quantile based Ensemble Model (SEQEM) is proposed to identify the spatial range that the station interactions take effects. (4) A comprehensive stochastic simulation framework is proposed to compare the user experience and system operations using different types of BSSs, which estimates actual origins-destinations of travel demands and integrates the user behavior model and the rebalance optimization model. The case-study results show that the user rerouting behaviors can indirectly affect system performances. Overall, systems with high usage-intensity can benefit from transitioning their station-based systems into hybrid systems.

In summary, this dissertation provides a holistic understanding of the mobility patterns, environmental benefits, and the impacts of system expansion and system types, which assists the policy making and the development of BSSs. The proposed models are transferable to different cities to support the development of sustainable micro-mobility systems.

1. INTRODUCTION

In this chapter, the basic components of bike share systems are first introduced in Section 1.1.1; then, Section 1.1.2 outlines the motivations and objectives of this dissertation. An overview of related research is provided in Section 1.2. The research gaps in existing studies and the corresponding contributions of this dissertation are summarized in Section 1.3. Lastly, Section 1.4 outlines the research scope and organization of this dissertation.

1.1 Background

1.1.1 Bike Share System (BSS)

Bike share is an emerging transportation mode in many countries, including the United States. Figure 1.1 shows that the annual trips taken from the bike share systems (BSSs) in the United States kept increasing from 2010 to 2019. Traditionally, the BSSs are station based, requiring the users to return the bikes to the docks at bike share stations (Fishman, 2016). As of 2019, 40 million trips were taken using station-based BSSs in the United States (NACTO, 2020). In recent years, dockless BSSs are enabled by the development of information and communication technologies (ICT). Dockless systems do not rely on docking stations and allow the users to park the bikes on the sidewalk. Because of this flexibility and the low capital investment required for a city to set up the system, the dockless BSSs are growing rapidly (Chen et al., 2020b; Dickey, 2019; Rinde, 2019). A third alternative, hybrid systems, have also emerged to integrate dockless and station-based BSS features. Hybrid systems allow cities that already have installed bike share stations to keep using the existing docking infrastructure but add the flexibility of parking bikes on-street (McMahon, 2019; Sisson, 2017). Both Chicago (The City of Chicago, 2019) and Philadelphia (City of Philadelphia, 2019) have launched pilot dockless bike share programs, temporarily testing the dockless operations in a small area, in addition to the current station-based systems.

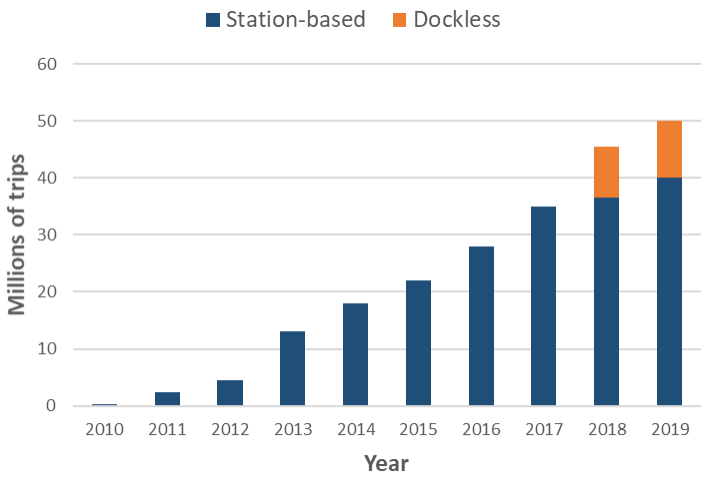


Figure 1.1. The growth of bike share ridership of station-based and dockless BSSs in the United States (NACTO, 2020)

A typical BSS consists of two major players (users and system operators) and three types of physical elements: bikes, smartphone applications, and docking stations (for station-based/hybrid systems). Their roles and relationships are described below:

- Users:

  Users ride the bikes, generate bike share trips and pay for the trips. There are mainly two types of users: subscribers and one-time users. Subscribers purchase a plan (e.g., monthly and weekly plans) and enjoy free rides if they use the bikes for a shorter time than a threshold (e.g., half an hour). Extra charges will be applied if the trip duration is longer than the threshold. In contrast, one-time users pay for each trip they make. Users make the decisions on whether to make the trip using bike share and thus are the demand sources of the system.

- System operators:

  The operators set up the systems and conduct the daily operations for a BSS. Their responsibilities include:

(1) Determining the location and size of docking stations in a station-based/hybrid system and controlling the number of bikes in each area. Sometimes the operators are only hired to operate the system, while the system design is determined by the city. For example,

New York City determines the location and size of the stations. The operator, Motivate, only supplies the bikes and operates the system (Institute for Transportation & Development Policy, 2018).

(2) Constituting the pricing scheme.

(3) Carrying out the maintenance of the systems, e.g., fixing the broken bikes.

(4) Rebalancing the bike fleets if needed. For instance, in a station-based system, if a station is full or empty, the users cannot return or check out a bike, and thus the operators may need to move some bikes to or from other stations. Similarly, a dockless BSS may also face the issue that some regions have too many bikes, while some other regions encounter bike shortages. Most of the operators use vehicles (e.g., trucks) to rebalance the bikes (Schuijbroek et al., 2017). Some use electric tricycles (e.g., in Oregon[1]). Some studies also proposed price-incentive methods for bike rebalancing (Fricker & Gast, 2016; Pan et al., 2019; Singla et al., 2015), which still requires more real-world practice to test its viability.

- Bikes:

  Some of the BSSs integrate Global Positioning System (GPS) devices on the bikes, which could track the locations of the bikes. In a station-based system, the users unlock a bike through a kiosk at the station. When a bike is returned to a docking station, the information for the ending of the trip is also recorded through the kiosk or the devices on bikes. In a dockless/hybrid BSS, the GPS devices on the bikes provide the location information, so both the users and system operators know where to find the bikes.

- Smartphone applications:

  Smartphone applications (apps) are key elements for a dockless/hybrid BSS. Users can view the locations of nearby available bikes on the apps and decide which bike to use. Apps are also used to unlock the bikes (e.g., by scanning the QR code on the bike). Although not

[1] Portland now using pedal-powered trikes to help rebalance bike share stations, https://bikeportland.org/2016/09/07/portland-now-using-pedal-powered-trikes-to-help-rebalance-bike-share-stations-191007, accessed on 2019/01/27.

necessarily, many station-based BSS operators also develop their apps, through which the users can view the status of docking stations (number of available bikes and docks).

- Docking stations:

  Docking stations only exist in station-based/hybrid systems. Such stations consist of a kiosk and many docks. The operators need to decide the number of docks in order to save the cost of station construction while also fulfilling the demands of users.

1.1.2 Motivations and objectives

For a new transportation mode such as bike share, it is fundamental and important to learn about its travel patterns, which helps understand the role that BSSs play in an urban transportation system. Specifically, the question "for what distance and duration do people consider using bike share?" needs to be answered. Such travel patterns affect the relationship between bike share and other transportation modes. On the one hand, bike share may compete with other transportation modes that have an overlap regarding the covered distance range. On the other hand, if bike share improves the transport efficiency within its covered distance range, it can enhance the multimodal transport connections for the whole transportation system (Shaheen et al., 2010).

Besides such potential benefits for multimodal transport connections, previous studies also claim that bike share can potentially benefit the cities and the society from multiple aspects: reducing public transit travel time (Jäppinen et al., 2013), alleviating traffic congestion (Faghih-Imani et al., 2017; Hamilton & Wichman, 2018), saving travel cost (Buehler & Hamre, 2015), and reducing greenhouse gas (GHG) emissions (Drynda, 2014; Shaheen et al., 2010). People may choose bike share for the benefits such as time saving and cost saving, while cities consider implementing a BSS for its overall benefits to the whole society. One of such overall benefits is the environmental benefit. One advantage of bike share is that it does not generate tailpipe emissions like cars and buses (Dave, 2010; Fishman et al., 2014). Therefore, bike share has great potential to improve the transportation sustainability. In this case, evaluating the environmental benefits can support the decision of launching BSSs and supporting the development of BSSs.

After launching a BSS in a city, the system operators may also make changes to the system to improve it. One of the most important changes is expanding the system. It is rare that a BSS is built at its full size when launching the system. Figure 1.2 shows the number of bike share stations in Chicago, New York, and Washington DC in different years. The station-based BSSs in the three cities were all first launched in year 2013. In the following years, the BSSs kept adding new stations to expand the systems. Such system expansions can improve the spatial service coverage of the BSSs and potentially increase the ridership (NACTO, 2020) and thus bring more benefits to the society.

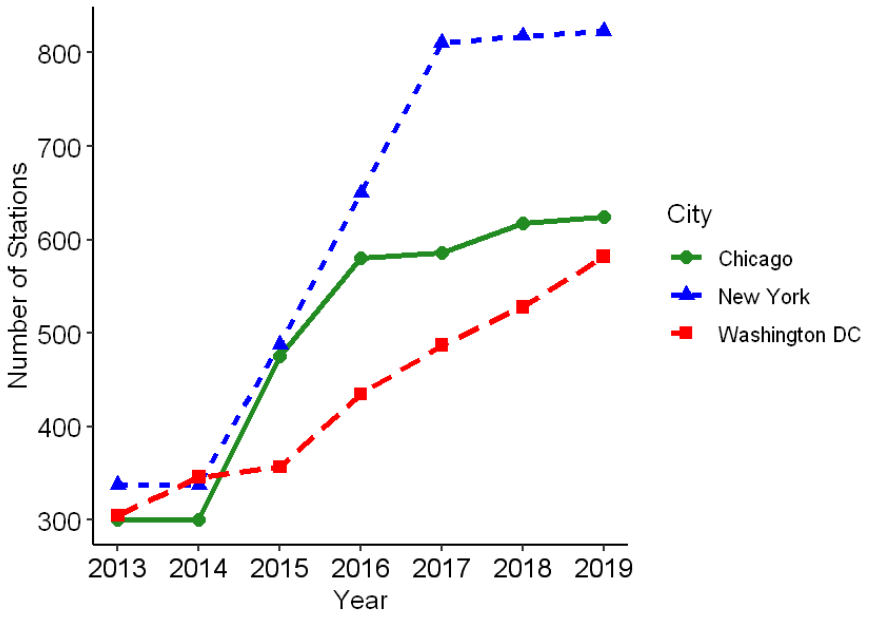


Figure 1.2. Number of stations in each year in Chicago, New York, and Washington DC

Besides system expansion, another important system change is changing the system types. A BSS could be implemented through different types: station-based, dockless, and hybrid. Different types of BSSs may affect user experience and may face different operational challenges. For the user experience, the bike accessibility (how easy it is to find a bike) could be different in different types of BSSs, because bikes are all located at the stations in station-based BSSs, while bikes are distributed more sparsely in dockless/hybrid BSSs. Also, depending on the different bike spatial distributions and bike parking requirements (users have to return bikes to stations in station-based BSSs, which is not required in dockless/hybrid BSSs), the users may spend different amounts of

excess time to pick up or return bikes in different types of BSSs. The convenience of bike pick-up/returning is important to users and thus impacts the ridership in the long-term. For operational challenges, bike rebalance in station-based BSSs needs to consider the station capacities. Therefore, the needs for bikes in some high-demand stations may not be met due to the capacity limits, which leads to unserved trips. In contrast, more bikes can be allocated to high-demand areas in dockless/hybrid systems. Nevertheless, dockless/hybrid systems face the issues that (1) the rebalance vehicles need to travel around to collect scattered bikes, and (2) the bikes parked on the sidewalks may clutter the streets. Therefore, comparing the system performances of different types of BSSs is critical for selecting the appropriate BSS types when launching a new system or improving an existing one.

Overall, motivated by the abovementioned aspects, this dissertation aims to improve understanding of travel patterns, environmental benefits, and the impacts of system changes (expanding the systems and changing system types) in BSSs to assist the policy making and development of BSSs.

1.2 Overview of related research

Previous studies about bike share mainly focus on the analysis of ridership (O'Brien et al., 2014; Zhou, 2015), the optimization of bike rebalance (Contardo et al., 2012; Fricker & Gast, 2016; García-Palomares et al., 2012), and the short-term demand prediction (Chai et al., 2018; Hulot et al., 2018; Jia et al., 2019; Li et al., 2016; Zhou et al., 2019). However, the existing studies about BSSs are limited in the understanding of mobility patterns (Section 1.2.1), environmental benefits (Section 1.2.2) and the impacts of system expansion (Section 1.2.3) and system type changes (Section 1.2.4). This section (Section 1.2) provides an overview of the related research and identifies the corresponding research gaps. More detailed literature reviews about different topics are provided in Chapters 2 – 5.

1.2.1 Bike share mobility patterns

The knowledge of the dynamics and statistical properties of human mobility can be applied to many fields, such as traffic forecasting (Jiang et al., 2009; Kitamura et al., 2000), urban planning (Horner & O'Kelly, 2001), epidemics (Grenfell et al., 2001; Hufnagel et al., 2004; Keeling et al.,

2001; Mari et al., 2012), genetics (Nash, 2012), and mobile network designing (Lee et al., 2009). For instance, understanding the statistical patterns of human movement can enable better simulation and design of the mobile networks of cellphones (Rhee et al., 2011), because wireless devices rely on the location of nearby nodes (e.g., base stations) to connect with the wider telephone networks. In recent years, researchers are able to better analyze human mobility patterns due to the ubiquitous availability of location tracking technologies such as Global Positioning System (GPS) and mobile devices (e.g., smartphones). Such studies have focused on the distributions of trip displacement (i.e., how far do people travel in one trip) and trip duration (i.e., how long does one trip take). For example, Brockmann et al. (2006) analyzed the circulation of bank notes and found that people's travel displacement follows a power law distribution and exhibits Lévy flight (random walk) behaviors.

Previous studies have analyzed the human mobility patterns for various transportation modes, but few paid attention to bike share trips. The majority of the previous studies focus on the mobility patterns of automobiles such as private cars and taxis (Bazzani et al., 2010; Cai et al., 2016; Liang et al., 2012; Veloso & Phithakkitnukoon, 2011). For instance, Jiang et al. (2009) examined the trajectories of taxi trips to study the mechanisms governing the Lévy flight behavior of human mobility and also compared the fitted results of power law distribution with those of lognormal distribution. They concluded that the Lévy flight behavior is driven by the underlying street network and the human mobility patterns can be reproduced by simulated random walks. In addition, some studies also paid attention to walking ((Koelbl & Helbing, 2003) and flights (Brockmann et al., 2006). Many studies focused on the long-distance trips and found that the trip displacement follows either power-law (Brockmann et al., 2006; Koelbl & Helbing, 2003; Rhee et al., 2011) or exponential (Liang et al., 2012; Yan et al., 2013) distribution. However, when the short-distance trips are also considered, some studies identified the best-fit distribution to be gamma distribution (Veloso & Phithakkitnukoon, 2011) and lognormal distribution (Jiang et al., 2009). Although existing studies have done extensive work to analyze human mobility patterns, very few researchers have analyzed the travel patterns of bike share trips. One major reason is that bike share is a relatively new mobility option and bike share trip data is only available in recent years. Analyzing the travel patterns could help the planning of transportation systems (Koelbl & Helbing, 2003) and provide critical information for the modeling of BSSs. For example, the

distribution of trip distance is important for modeling the bike flows when simulating a BSS (Chen et al., 2020b). Therefore, the first research gap is identified as the lack of a comprehensive analysis on the travel patterns of bike share trips (**Gap 1**).

1.2.2 Environmental Benefits of BSSs

Among the many potential benefits of BSSs, the environmental benefit is one of the major factors that motivate the decision makers to introduce BSSs into the cities. Quantifying the environmental benefits of BSSs can help the city decision makers understand the role that a BSS plays in the whole urban transportation systems, and thus make the policies and BSS development plans accordingly. Understanding the modal shift is important to evaluate the environmental benefits of BSS. Whether bike share reduces transportation energy use and emissions depends on the transportation mode replaced by the bike share trips. If bike share trips replace personal vehicle or taxi trips, the bike share systems will reduce transportation fuel use and emissions. However, if the bike share trips replace walking trips, they may actually increase transportation energy use, due to the energy consumption required to build the stations, manufacture the bikes, and operate the system (e.g., transporting the bikes between stations to rebalance the bike availability). Therefore, in order to estimate the environmental benefits of bike share systems, it is critical to consider the transportation mode substitution by bike share trips. In existing literatures, there are limited studies trying to quantify the environmental benefits of BSSs, in which the mode substitution is either based on simplified assumptions (Anderson, 2015; Zhang & Mi, 2018) or shifted concepts (e.g., using percentage of users information from bike share user surveys as percentage of travelled miles values (Fishman et al., 2014)). The environmental benefits estimated using these methods may lead to biased result. Therefore, better methods that consider the heterogeneity of bike share trips are needed to better quantify the environmental impacts of BSSs (**Gap 2**).

1.2.3 BSS expansion

In the process of BSS expansion, some studies have observed that adding new stations can impact the demand of existing stations. Wang and Lindsey (Wang & Lindsey, 2019) applied a difference-in-difference analysis on the bike share usage of members (long-term users). They found that

members were more likely to increase their bike share usage when the distance from their homes to the nearest stations were shortened. Zhang et al. (2016) found that adding new stations can attract new users and encourage the usage of long-term users, but new stations could also compete with old stations. However, they did not point out in what conditions that the stations compete with each other. Using mixed-effect regression model, Hyland et al. (2018) revealed that the demand of existing stations would decrease when the number of stations within 0.5 miles increased, while the demand would increase if the number of stations within 0.5-3.1 miles increased. While this study brings insights on station interactions, it also has two limitations: first, the distance threshold of 0.5 miles is selected arbitrarily; second, the study fails to answer whether the impact is homogeneous for any distance within the range of 0-0.5 miles (or 0.5-3.1 miles). For example, a new station that is 1 mile away could impact the demand of an existing station to a different extent compared with a new station that is 2 miles away. Therefore, a more detailed analysis about such station interactions is needed, which is important for the designing of a balanced station network, i.e., a station network that covers sufficient service areas but also avoids the competitions between stations that are too close). This corresponds to the third gap: there lacks a detailed understanding of the station interactions in the bike share station networks (**Gap 3**)

Additionally, such station interactions have not yet been considered in the existing demand prediction models for BSS expansion. Demand estimation for BSS expansion helps the decision makers to evaluate the feasibility of the expansion plan and determine the capacity for the new stations. Compared with plenty of literature about the short-term prediction of BSS demand which predicts the demand for a certain station based on its historical trips (Chai et al., 2018; Jia et al., 2019; Lin et al., n.d.; Wang & Kim, 2018), the literature about the demand prediction for BSS expansion is very limited. One distinction between the short-term prediction and system expansion prediction is that there is no historical trip record of new stations for BSS expansion. Existing studies highly rely on external information to make the predictions (Liu et al., 2017; Watson & Telenko, 2019), such as data about social-demographic (e.g., population density) and points of interest (POI) (e.g., the number of restaurants) around a certain station. It requires time and effort to collect such external data; also, different studies used different categories of external data, making it hard to generalize the models to other cities. Additionally, these studies have not accounted for the station interactions. They only report the prediction performance on new stations,

but ignore that the new stations could also impact the demand of existing stations. In the case that a new station complements the demands of its nearby existing stations, if the demands of existing stations are assumed to remain the same, the total demands of the whole system will be under-estimated, which may lead to wrong decisions on the expansion plan. In summary, existing studies did not consider the station interactions to predict the demands for BSS expansion (**Gap 4**).

1.2.4 System types

As mentioned in Section 1.1.1, BSSs have multiple types: station-based, dockless, and hybrid. So far, when evaluating or improving the system performance, the majority of existing studies focuses on only one specific type of BSSs, instead of comparing different types of BSSs. Most of such studies are about station-based BSSs (Alvarez-Valdes et al., 2016; Brinkmann et al., 2015; Hulot et al., 2018; Lin et al., n.d.; Médard de Chardon & Caruso, 2015), because station-based BSSs have a longer history and have released more trip data to the public. In recent years, a growing body of literature has paid attention to dockless BSSs (Caggiani et al., 2018; Liu et al., 2018; Xu et al., 2018a). Due to the fact that cities have considered hybrid operations only since very recently (City of Philadelphia, 2019; The City of Chicago, 2019), studies about hybrid BSSs are very limited (Albiński et al., 2018).

Comparing the performances of different types of BSSs can assist in selecting the appropriate system types to better take advantage of the benefits of BSSs. However, very few studies have quantitatively compared the differences from the perspectives of users and system operators. The few studies discussing different types of BSSs are mainly in two groups: (1) macro-scope analysis comparing the bike share use, policy, and business models across multiple cities (Gu, Kim, & Currie, 2019; Hirsch et al., 2019; Susan et al., 2014) and (2) analysis of the different usage patterns when both station-based and dockless BSSs are available in the same city (M. Chen, Wang, Sun, Waygood, & Yang, 2020; Ji, Ma, He, Jin, & Yuan, 2020; Lazarus, Pourquier, Feng, Hammel, & Shaheen, 2020; H. Li, Zhang, Ding, & Ren, 2019). However, the observed differences about ridership and the usage patterns from these studies are also the results of many external factors, such as population density, pricing schemes, and the assignment of service areas, which are not tied to a specific type of BSSs. Therefore, the results from these studies cannot conclude which type of BSSs is better for a specific city regarding the user experience and operational challenges.

In summary, there is a critical need for models that can quantify system performances of different types of BSSs from the perspectives of both users and operators when serving the same demands (**Gap 5**).

1.3 Research Gaps and Contributions

This section summarizes the identified research gaps in Section 1.2 and discusses the corresponding contributions of this dissertation to address these gaps.

- **Gap 1**: The lack of comprehensive analysis on the travel patterns of bike share trips.

  This dissertation analyzes the statistical property of trip distance and duration for bike share trips. To provide a thorough understanding of bike share travel patterns, both short trips and long trips, as well as trips with different purposes, are analyzed. Trip data in eight cities (Seattle, Los Angeles, Bay Area, Philadelphia, Boston, Washington DC, Chicago, and New York) is collected and analyzed to compare the bike share travel patterns in BSSs of different system sizes. The insights gained in this study could help the design of bike share station networks.

- **Gap 2**: Existing studies evaluating the environmental impacts of BSSs neglects the impacts of trip purpose, trip distance, start time, and heterogeneous trip mode substitution.

  This dissertation proposes a Bike Share Emission Reduction Estimation Model (BS-EREM) to quantify the environmental benefits of BSSs. The BS-EREM model stochastically estimates the transportation modes substituted by bike share trips, considering the detailed trip information and heterogeneous transportation mode substitution. The model improves upon existing literature through the following aspects: (1) the mode substitution estimation takes account of trip distance, trip purpose, start time, and the public transit accessibility near bike share stations, which provides more accurate estimation of the environmental benefits of bike share trips. (2) The BS-EREM relies on National Household Travel Survey (NHTS) and bike share trip data to estimate the environmental benefits. Therefore, it can

be directly applied to all cities that have such data. In this dissertation, the BS-EREM is applied to eight cities in the United States, to learn about the environmental benefits of BSSs with varying system sizes. (3) The case study also provides detailed analysis such as unit emission reduction per trip and per mile travelled, as well as the spatial distribution of the environmental benefits, which could help the station siting and the planning of BSSs.

- **Gap 3**: There lacks a detailed understanding of station interactions in the bike share station networks.
- **Gap 4**: Existing studies did not consider the station interactions to predict the demands for BSS expansion.

To address these gaps, this dissertation applies a piecewise regression to formulate the relationship between the demand changes of existing stations and their distance to the nearest new stations, which improves the understanding of the station interactions and brings insights on station network designing. This dissertation also proves that the spatial station network structures, represented by fine-grained concentric station densities, can improve the predictions for BSS expansion. A Spatial Eccentricity Quantile based Ensemble Model (SEQEM) is proposed. It requires no external socio-demographic and point-of-interest data but outperforms models using the external information. Overall, the insights and the models in this dissertation can improve the planning for system expansion.

Gap 5: There is a critical need for models that can quantify system performances of different types of BSSs from the perspectives of both users and operators when serving the same demands.

This dissertation proposes a stochastic simulation framework to compare how different system types may impact user experience and rebalance needs. The proposed framework estimates actual origins-destinations of travel demands and integrates a user behavior model and a rebalance optimization model. The model is applied to three case study cities — Los Angeles, Philadelphia, and Chicago — to demonstrate how the simulation results

of different types of BSSs can shed light on BSS improvement and guide potential system type changes.

1.4 Research Scope and Organization of the Dissertation

The majority of this dissertation (Chapter 2 – 4) focuses on the station-based BSSs. The reasons are as follows:

(1) Station-based BSS is the dominating types of BSSs. In 2019, 80% of the bike share trips in the U.S. are generated in station-based systems (NACTO, 2020).

(2) More trip data are available for station-based bike share systems, which allows better understanding and more detailed modeling of station-based BSSs. The trip data of station-based BSSs collected in this dissertation is from as early as year 2011, while most of dockless BSSs are introduced in recent 2-3 years. Besides, most of the dockless BSS companies do not release their trip data to the public, which makes it more difficult to study dockless BSSs.

(3) System expansion is a more critical issue for station-based BSS operators, because the installation of new stations incurs high initial expense and the service areas are more restricted by the station locations.

Chapter 5 compares different types of BSSs. Although the input data is still from station-based BSSs, this dissertation proposes a method to redistribute the bike share demands to wider spatial areas, which enables the modeling of dockless and hybrid BSSs.

The framework of the dissertation is depicted in Figure 1.3. The dissertation can be divided into three parts; and the corresponding research questions are listed below:

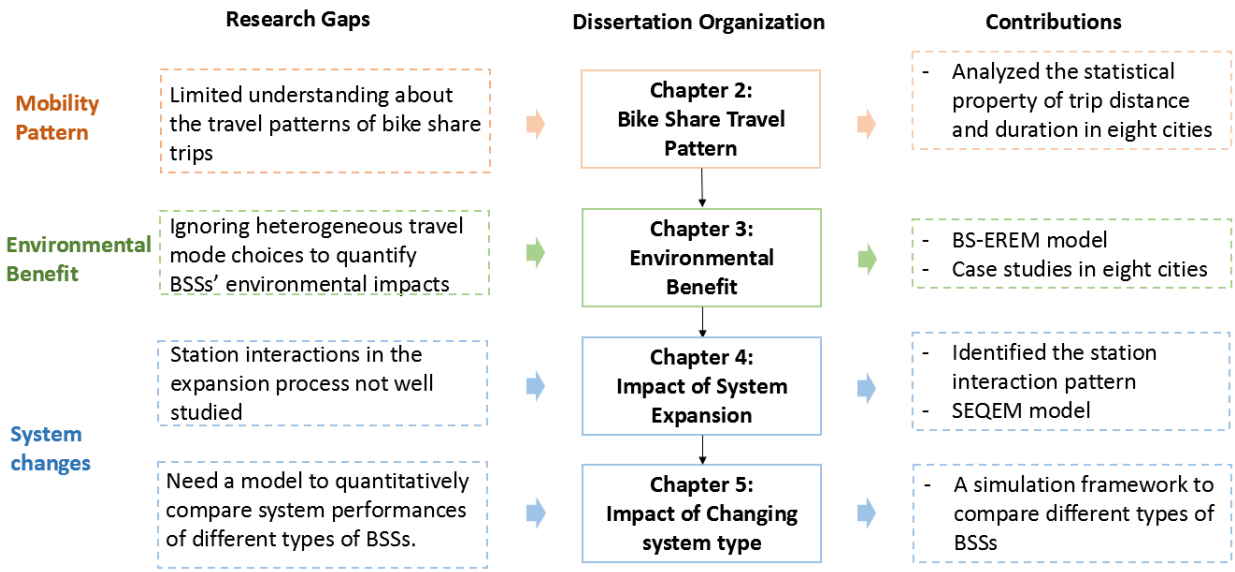


Figure 1.3 Framework of the dissertation

Part 1: Understanding the travel pattern of bike share trips (Chapter 2)

- **Q1**: which statistical model best fits the distributions of the distance and duration of bike share trips?

Part 2: Quantifying the envonmental benefits from bike share trips (Chapter 3)

- **Q2**: what transportation modes do bike share trips replace?
- **Q3**: how much do BSSs contribute to the greenhouse gas emission reductions?

Part 3: Modeling the system changes. This part can be partitioned into two sub-sections:

(1) Analyzing the station interactions in system expansion and their applications to demand prediction (Chapter 4)

- **Q4**: how the station interactions change with the distance between new and existing stations?
- **Q5**: how to use the information of station interactions for BSS expansion demand prediction?

(2) Comparing different types of BSSs (Chapter 5)

- **Q6**: which types of BSSs have better performances regarding user experience and system operations when serving the same demands?

Last, Chapter 6 summarizes the key findings of this dissertation and discusses the directions for future research.

2. UNDERSTANDING THE TRAVEL PATTERN OF BIKE SHARE

A modified version of this chapter has been published in a journal paper: Kou, Z., & Cai, H. (2019). Understanding bike sharing travel patterns: An analysis of trip data from eight cities. *Physica A: Statistical Mechanics and its Applications*, 515, 785-797.

2.1 Introduction

The knowledge of statistical properties of human mobility can be applied to many fields, including urban planning (Horner & O'Kelly, 2001), epidemics (Grenfell et al., 2001; Hufnagel et al., 2004; Keeling et al., 2001; Mari et al., 2012), genetics (Nash, 2012), and traffic forecasting (Jiang et al., 2009; Kitamura et al., 2000). Researcher have analyzed the distributions of the displacement and duration of human travel records using data from various sources. Brockmann et al. (2006) analyzed the records of bank notes and revealed that people's travel displacement follows a power law distribution and exhibits Lévy flight (random walk) behaviors. This Lévy flight pattern has been previously observed in animal movements (Atkinson et al., 2002; Ramos-Fernández et al., 2004; Viswanathan et al., 1996) and further been confirmed for human mobility by Rhee et al. (2011) using GPS traces of 44 volunteers in different outdoor settings. Based on cell phone traces, González et al. (González et al., 2008) found that, although the distribution of displacement for human motion follows a truncated power-law, Lévy flight isn't sufficient to explain the spatial and temporal regularities shown after the correction of the inherent anisotropy of each individual's trajectory and the different scale of travel distance. Using the same data set, Song et al. (Song et al., 2010) further proposed a mechanism, a combination of exploration and preferential return, to explain the human mobility scaling properties and regularities.

Researchers reported that trips taken using different transportation modes may have different mobility patterns. Yan et al. (Yan et al., 2013) pointed out that the displacements by a single transportation mode exhibit an exponential distribution, instead of the power law distribution previously reported, due to the influence of the traveling cost. Kolb and Helbing (Koelbl & Helbing, 2003) showed that the scaling laws in human travel behavior is related to the different energy consumption required for physical activities when travelling in different transportation modes

(walking, car, bus, and train). Higher energy consumption rate from physical activities will reduce the average travel times. Liang et al. (Liang et al., 2012) analyzed the taxi trips in Beijing and found that both the displacement and trip duration of taxi trips can be better fitted by an exponential distribution instead of a power-law distribution. However, in the analysis of taxi GPS traces in Lisbon, Portugal, Veloso and Phithakkitnukoon (2011) found that trip distance can be represented with a gamma distribution. While these studies focus on the mobility of the taxi passengers (considering only trips when the taxis are occupied by passengers), Cai et al. (Cai et al., 2016) further pointed out that the travel patterns of taxis, including both occupied and unoccupied trips (passenger searching or vehicle relocation), can be better fitted by the truncated power-law distribution for shorter trips and by the exponential distribution for trips that are longer than 30 miles. Bazzani et al. (Bazzani et al., 2010) studied the GPS data of private cars in Florence, Italy and found that the single-trip length follows an exponential behavior in short distance scale but favors a power law distribution for trips longer than 30 km. Overall, trips made by vehicles (taxi or private car) are most frequently analyzed in recent studies. Because the majority of the car trips are in long distance, short trips (e.g., less than one mile) are often neglected. When excluding the short trips, existing studies show that the trip displacement follows either power-law (Brockmann et al., 2006; Koelbl & Helbing, 2003; Rhee et al., 2011) or exponential (Liang et al., 2012; Yan et al., 2013) distribution. On the other hand, when short trips are included in the analysis, gamma distribution (Veloso & Phithakkitnukoon, 2011) and lognormal distribution (Jiang et al., 2009) may provide better fit.

Most of the previous studies focus on the travel patterns of trips taken by automobiles (e.g., taxis and private cars) because of the data availability of large-scale trip trajectory data recorded by GPS devices. Researchers have also analyzed other transportation modes such as walking (Koelbl & Helbing, 2003) and flights (Brockmann et al., 2006) with travel survey or bank notes data, which has lower data accuracy on the trip information. Few studies paid attention to the mobility patterns of trips taken by bikes. Part of this lack of understanding is due to the lack of bike trip data. Kolb and Helbing [18] analyzed the relationship between the energy consumption rates (from physical activities) and the travel time of different transportation modes (including biking). However, the statistical pattern of biking trips was not analyzed in this study. As a critical component of the multi-modal transportation systems, bike trips are playing an increasingly important role in urban

mobility. As one type of "active travel", biking not only can contribute to the sustainability of urban transportation by reducing traffic congestion and emissions, but also can help improve human health (Bullock et al., 2017). With the emerging sharing economy, more and more cities are implementing bike share programs, increasing the convenience of traveling by bikes in the city. As of January 2017, more than 100 cities in the United States have developed bike-sharing programs (Malouff, 2017). Additionally, the trip data from several bike share programs are made publically available, providing an unprecedented opportunity to understand the human mobility using shared bikes. Such knowledge will provide fundamental basis for researchers to model the use of bike share and the associated multi-modal transportation systems, inform bike share system design and operation, and guide policy decisions for sustainable transportation development.

Previous studies about bike share mainly focus on the analysis of ridership (Mateo-Babiano et al., 2016; O'Brien et al., 2014; Ome & Latifa, 2014; Zhou, 2015) and the optimization of system operation (Brinkmann et al., 2015; Chemla et al., 2013; Contardo et al., 2012; Fricker & Gast, 2016; García-Palomares et al., 2012; Li et al., 2015), few studies focus on the statistical patterns of bike share trips. Jurdak (Jurdak, 2013) studied the bike share trips in Boston and Washington, D.C. in the United States, and concludes that the distribution of trip durations in both cities fit a power law distribution for long trips (longer than an hour). However, the distribution of long trips do not reflect the travel patterns of bike share trips, because the majority of the bike share trips are short (this will be discussed in Section 2.2 of this chapter). Therefore, to fill the gap of understanding the travel patterns of bike share trips, this research analyzed the distribution of both trip distance and trip duration for bike share trips collected in eight cities in the U.S. The major contributions of this study include: 1) the analysis includes both short trips and long trips (generally, a threshold is set where the distance/duration reaches the highest probability (density) in the probability density plot; then trips with distance/duration that is larger than the threshold will be treated as long trips; for example, the threshold for duration varies from 5 to 13 minutes depending on the different cities and different trip purposes); 2) this study compares the travel patterns of bike share trips with different trip purposes (commuting versus touristic trips); and 3) this study compares the travel patterns of bike share in eight cities, which include bike share systems of different sizes and operation time.

2.2 Data and Methods

2.2.1 Data

This study analyzes the bike share trip data from eight programs located in eight different cities in the United States: Seattle, Los Angeles, Bay Area, Philadelphia, Boston, Washington D.C., Chicago and New York. The bike-sharing programs in these cities are station-based systems, in which a bike has to be checked out from and returned to a docking station. For each trip, the time and station is recorded when the bike is checked out and returned. The locations of the stations can be mapped with the station list, in which the longitude and latitude of each station is provided. Table A. 1 in the APPENDIX A shows a sample of the data from Pronto, the bike share program in Seattle. Data for other cities contains similar information. The bike share programs in different cities are launched at different time and have different scales (in terms of the number of stations and bikes). Table 2.1 shows an overview of the bike share programs in each city. This study analyzed all trips taken in 2016 to study the travel patterns of bike share. To indicate the scale of spatial coverage of each program, the system diameter for each program is measured, which is defined as the longest Euclidian distance between any two stations (the stations list kept changing with the system expansion over time and the station list at the end of year 2016 is used in this study). Based on the ridership, the total number of stations, and the system diameter, the programs can be naturally separated into two groups, smaller systems (those in Seattle, Los Angeles, Bay Area, and Philadelphia) and larger systems (those in Boston, Washington DC, Chicago, and New York). One thing to note is that Bay Area's bike share system contains multiple sub-systems in several cities, thus only intra-city trips in San Francisco are included (accounts for 89.7% of the trips) in this study to make it comparable to other cities (more information on this is provided in Section A.2 of APPENDIX A).

Table 2.1. Basic information of the eight bike share programs

City	Seattle	Los Angeles	Bay Area	Philadelphia	Boston	Washington, DC	Chicago	New York
Program Name	Pronto	Metro	Ford Gobike	Indego	Hubway	Capital	Divvy	Citi
Program Launch Date	Oct., 2014	June, 2016	Sep., 2013	July, 2013	Dec., 2011	Jan., 2014	July, 2013	July, 2013
Total Number of trips in year 2016	102,606	184,345	210,494	499,306	1,236,199	2,562,718	3,595,383	10,262,649
Number of Stations as of Dec. 2016	59	64	74	119	327	407	581	687
System diameter (miles)	4.71	2.85	2.33	5.06	8.57	14.28	23.29	11.17

Note: the programs are listed in the order of the program size (the number of stations) from the smallest (left) to the largest (right)

2.2.2 Estimation of trip distance, duration, and purpose

Because the data only include trip origins (check-out station) and destinations (return station), the actual trajectory of each trip is not available. Therefore, the trip distance has to be estimated using the locations of the trip start and end stations. Considering the fact that the movement of bikes is likely restricted by the street network and not all roads (e.g., highways) can be traveled by bikes, to more accurately estimate the trip distance, the Google Maps Distance Matrix API is used to calculate the trip distance and duration between two stations, specifying biking as the transportation mode. This distance can better reflect the actual bike travel along the pathways suitable for biking. This study also tested how trip start time can impact the trip distance and duration but did not observe significant differences (Google Maps Distance Matrix API only considers the real-time road traffic conditions for driving, not biking). Considering that bike travels are less impacted by road conditions (e.g., congestion), which is one of the reasons some people prefer biking to driving (Alter, 2008; Lobo, 2013), it is assumed that the travel distance and duration between the same pair of stations would stay the same regardless of the travel time. Because there is no sufficient information to estimate the actual traveling distance of round trips,

for which the bikes were checked out from and returned to the same station, round trips (consisting of 2% to 10% of the trips as shown in Table 2.2) are excluded from the analysis.

While many previous studies used trip displacement (e.g., the Great Circle Distance between the origin and destination) to measure the human movements (González et al., 2008; Rhee et al., 2011), the actual distances traveled within the urban street network can better reflect human mobility using bikes. As shown in Figure 2.1, the distance estimated by Google Maps API is on average 1.5 times of the great circle distance, while the ratio can be as high as 8 times.

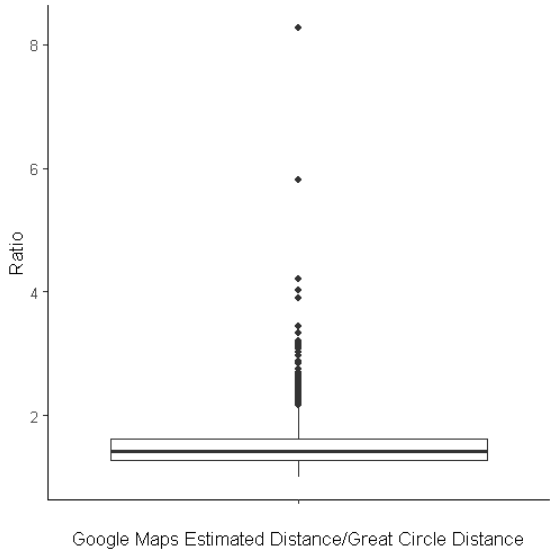


Figure 2.1. Box plot of the ratio of Google Maps estimated distance to the trip displacement measured by the great circle distance (using data from Seattle as an example)

Bike share systems are mainly used for commuting and touristic trips (LDA Consulting, 2016). The two types of trips will have very different trip duration patterns: commuters will bike directly from the start station to the end station while the tourists tend to stop at different points-of-interest locations along the way. As a result, the actual trip duration measured by the difference between trip end (bike return) and start (bike check-out) time can well reflect the traveling time for the commuting trips but will be much longer than the traveling time for the touristic trips. Because the original trip data do not contain trip purpose information, this work compares the actual trip duration (the difference between trip start and end) to the expected trip duration (estimated by Google Maps API) to estimate the trip purpose. As shown in Figure 2.2 a, the trip duration

estimated by Google Maps API is linear to the estimated distance, while the actual trip duration could be much longer than the estimated one. The distribution of the actual trip duration has a longer tail for long trips (e.g., longer than 15 minutes for Seattle) than that of the estimated duration (Figure 2.2 b). This means that many bike share users did not ride from the origin directly to the destination through the shortest route. Therefore, this study defines the trips whose duration is within 1.3 times of the duration estimated by the Google Maps API as the commuting trips and trips taken longer than that as touristic trips. The Google Maps API assumes a baseline speed of 10 miles/hour for biking (Lobo, 2014). This cutoff line of 1.3 is arbitrarily selected, considering a 30% buffering for possible delays (e.g., due to bike adjustment, necessary detours, lower biking speed) compared to the expected duration estimated by the Google Maps API. A sensitivity analysis is conducted to evaluate the impact of this cutoff line and is discussed in Section 2.3.3. The commuting trips make up about 59% to 76% of the total trips in the eight cities (Table 2.2) and are more concentrated in the shorter range (Figure 2.3).

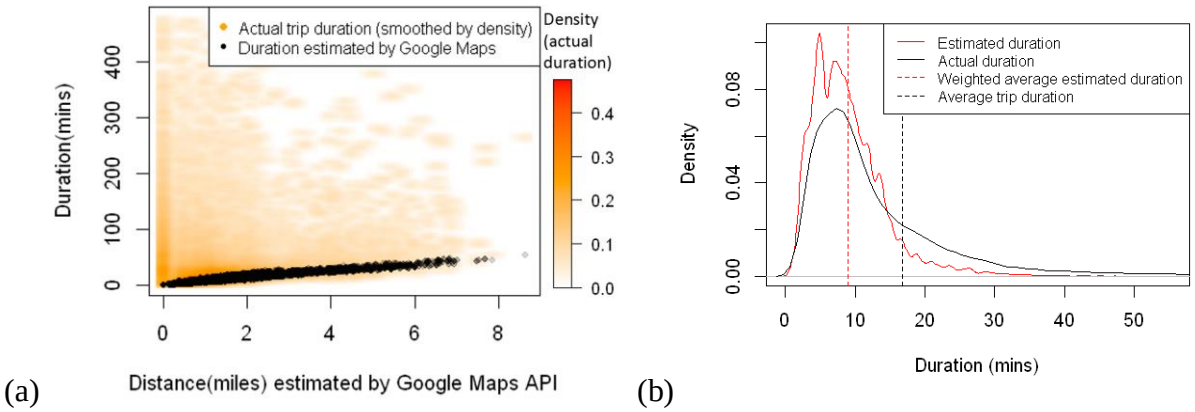


Figure 2.2. Trip duration in Seattle: (a) Comparison of the relationship between the trip duration and the estimated distance for the actual and estimated trip duration and, (b) Trip duration density plot (the total trip between each origin-destination pair is used as the weight for the weighted average estimated duration calculation)

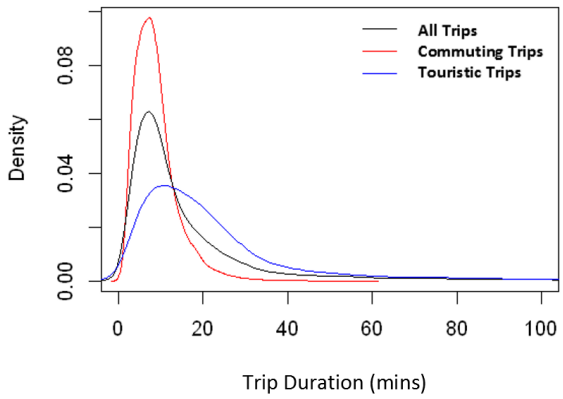


Figure 2.3. Probability density plot of trip duration for commuting trips and touristic trips in Seattle

Table 2.2. Basic statistics of the trip distance and duration

City			Seattle	Los Angeles	Bay Area	Philadelphia	Boston	Washington, DC	Chicago	New York
Percentage of round trips			8.40%	10.60%	2.60%	8.10%	3.10%	4.60%	4.20%	2.06%
Percentage of commuting trips			58.43%	63.61%	75.55%	68.59%	68.98%	58.56%	60.78%	62.32%
Commuting trips	Distance (miles)	Maximum	7.85	3.14	3.37	7.30	8.77	27.21	23.77	13.88
		Median	1.16	0.91	1.23	1.39	1.46	1.43	1.39	1.31
		Average	1.27	1.02	1.24	1.56	1.70	1.72	1.71	1.68
		Standard Deviation	0.74	0.50	0.53	0.79	1.02	1.14	1.14	1.21
	Duration (minutes)	Maximum	58.86	28.00	28.45	57.00	64.07	158.28	148.63	82.68
		Median	7.89	4.00	8.15	10.00	9.28	9.47	9.33	8.80
		Average	8.84	4.57	8.55	11.02	10.79	11.10	10.98	10.85
		Standard Deviation	5.11	4.91	3.65	5.75	6.28	6.94	6.75	7.28
Touristic trips	Distance (miles)	Maximum	8.65	3.29	3.37	7.20	10.60	25.34	18.28	14.16
		Median	0.95	0.84	0.93	1.22	1.19	1.16	1.26	1.21
		Average	1.15	0.94	1.05	1.37	1.42	1.40	1.55	1.48
		Standard Deviation	0.85	0.50	0.54	0.79	0.92	1.03	1.13	1.01
	Duration (minutes)	Maximum	479.66	719.00	716.70	719.00	719.37	719.72	719.85	719.98
		Median	17.39	14.00	11.45	20.00	16.42	16.40	17.03	15.30
		Average	27.97	23.62	21.20	31.58	22.41	26.01	21.89	20.09
		Standard Deviation	39.33	44.66	43.98	50.49	33.01	37.60	26.78	27.09

2.2.3 Methods for fitting the distributions of trip distance and duration

Short trips that are less than one mile are often neglected in existing studies analyzing travel patterns (for example, in (Cai et al., 2016; Liang et al., 2012)). However, because the majority of the bike share trips are relatively short (the median distance for both commuting and touristic trips is less than 1.5 miles, as shown in Table 2.2), it is critical that these short-distance trips are included in the analysis. Therefore, this work evaluates the travel patterns both for all trips (including the short ones) and for long trips only (to allow comparison between this study and those focus on longer trips). Based on the shape of the density plot (Figure 2.3), this study first selected three common right-skewed distributions, Weibull, gamma, and lognormal as candidate distributions to fit the entire data (including both short and long trips). The Lévy flight nature of human travelling is also examined by fitting the Power Law and exponential distribution to the right-side tail (long trips). This study then fits the trip data to the selected distributions, using the maximum likelihood estimation (MLE). To identify the goodness of fit, this study conducts a Kolmogorov-Smirnov (K-S) test (D'Agostino RB, 1986) and considers a K-S statistics value lower than 0.05 as acceptable for a good fit. If the K-S test considers multiple distributions as acceptable, this work further uses the Akaike information criterion (AIC) and Bayesian information criterion (BIC) methods (Johnson & Omland, 2004) for model selection. The distribution with the lowest AIC and BIC is selected as the best fit (Posada & Buckley, 2004).

Table 2.3 Probability density function of the models considered in this study (x > 0)

Model	Equation	Parameters
Weibull	$f(x) = \frac{k}{\lambda}(\frac{k}{\lambda})^{k-1} e^{-(x/\lambda)^k}$	Shape parameter k and scale parameter λ
Gamma	$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$ where $\Gamma(\alpha)$ is a complete gamma function	Shape parameter α and scale (rate) parameter β
Lognormal	$f(x) = \frac{1}{x} \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(\ln x - \mu)^2}{2\sigma^2})$	Mean (log) μ and standard deviation (log) σ
Power Law	$f(x) = ax^{-k}$	Exponent k
Exponential	$f(x) = \lambda e^{-\lambda x}$	Rate parameter λ

2.3 Results and Discussions

2.3.1 The distribution of trip distance and duration for all trips (short and long)

This study fitted the trip distance and duration for the commuting and touristic trips separately with three common right-skewed distributions: lognormal, gamma, and Weibull (see Table 2.3 for the equations of these distributions). According to the fitted results (Table 2.4), the eight systems do not have a common distribution pattern regarding trip distance for the commuting trips. For most of the systems except for Bay Area, both gamma and lognormal distributions provide a good fit with K-S statistics below the 5% significance level. Based on the AIC/BIC values, the best fit for the large systems (Boston, Washington DC, Chicago, New York) is lognormal distribution, while gamma distribution is a better fit for the smaller systems (Seattle, Los Angeles, and Philadelphia) except for Bay Area which fits better with the Weibull distribution. The details of the goodness of fit statistics for tourist trip distance, commuting trip duration, and tourist trip duration are listed in the APPENDIX A, Table A. 3 to A. 5, respectively.

Table 2.4. Results of model selection for commuting trip distance

			Seattle	Los Angeles	Bay Area	Philadelphia	Boston	Washington, DC	Chicago	New York
Weibull	Parameters	**Shape**	1.8512	2.1665	2.5134	2.0996	1.7896	1.6420	1.6144	1.5232
		Scale	1.4361	1.1499	1.4011	1.7712	1.9261	1.9317	1.9218	1.8760
	Goodness-of-fit statistics/criteria	**K-S statistic**	**0.0658**	**0.0513**	**0.0497***	0.0555	0.0548	0.0519	0.0547	0.0674
		AIC	107316.3	137674.1	215312.1	693533.8	2161993.0	3880947.0	5717157.5	17356522.0
		BIC	107334.1	137693.2	215331.8	693555.1	2162016.3	3880971.3	5717182.6	17356549.0
Gamma	Parameters	**Shape**	3.5464	4.0595	4.9903	4.1294	3.0170	2.7110	2.5535	2.3567
		Scale	2.7932	3.9995	4.0206	2.6417	1.7713	1.5799	1.4937	1.4054
	Goodness-of-fit statistics/criteria	**K-S statistic**	**0.0382***	**0.0304***	0.0682	**0.0284***	**0.0472***	**0.0354***	**0.0462***	0.0570
		AIC	101513.2	134402.0	216873.8	671542.5	2113945.2	3771463.5	5585434.3	16907864.0
		BIC	101531.1	134421.1	216893.5	671563.9	2113968.4	3771487.9	5585459.4	16907892.0
Lognormal	Parameters	**meanlog**	0.0912	-0.1133	0.1126	0.3208	0.3578	0.3444	0.3278	0.2899
		sdlog	0.5500	0.5264	0.4791	0.5136	0.6046	0.6345	0.6563	0.6783
	Goodness-of-fit statistics/criteria	**K-S statistic**	0.0504	**0.0384***	0.0816	**0.0205***	**0.0237***	**0.0152***	**0.0176***	**0.0131***
		AIC	100235.9	138111.9	226762.0	673148.5	2104083.5	3722622.1	5514671.9	16539126.0
		BIC	100253.8	138131.0	226781.7	673169.8	2104106.7	3722646.5	5514697.0	16539154.0
Best Fit Distribution			Gamma	Gamma	Weibull	Gamma	Lognormal	Lognormal	Lognormal	Lognormal
Parameters			$\alpha = 3.55$ $\beta = 2.79$	$\alpha = 4.06$ $\beta = 4.00$	$k = 2.51$ $\lambda = 1.40$	$\alpha = 4.13$ $\beta = 2.64$	$\mu = 0.36$ $\sigma = 0.60$	$\mu = 0.34$ $\sigma = 0.63$	$\mu = 0.33$ $\sigma = 0.66$	$\mu = 0.29$ $\sigma = 0.68$

Notes: 1. * Significant at 0.05.

2. Model parameters corresponds to the parameters in Table 3.

Table 2.5 summarizes the best fit models as well as the parameters for all cities and all cases. In general, lognormal distribution provides the best fit for cities with larger bike share system (having a diameter larger than or around 10 miles), for both trip distance and distribution. The mean μ of lognormal distribution is between 0.29 and 0.36 miles for commuting trips and between 0.12 and 0.21 miles for touristic trips. Tourists tend to return the bikes in areas near the trip origins, and thus have shorter average trip "distance" than commuters (note that the distance here better reflect the real distance travelled by commuting trips than touristic trips). Regarding trip duration, on the contrary, tourists will travel longer time (mean μ is around 2.75 minutes for lognormal distribution) than commuters (mean μ is around 2.2 minutes), as expected. For cities with smaller system (whose diameter is less than 8 miles and have less than 100 stations), gamma distribution provides the best fit to describe the trip distance in most of the cases with some exceptions (the distance of commuting trips in Bay Area fits better with Weibull distribution; the distance of touristic trips in Seattle fits better with lognormal distribution). For trip duration, none of the selected distributions can fit many of the smaller systems with a significant K-S statistic. However, if the significant level of the K-S statistics is extended from 0.05 to 0.1, the lognormal distribution can be viewed as providing a good fit for the duration of touristic trips in all cities (Table A. 5 in the APPENDIX A). If such significant level extension is also applied to the trip distance analysis, the model selection will remain the same except that lognormal will be preferred for commuting trip distance.

Table 2.5. Parameters of the best fit models

		Seattle	Los Angeles	Bay Area	Philadelphia	Boston	Washington, DC	Chicago	New York
Commuting trip distance	Best Fit Distribution	Gamma	Gamma	Weibull	Gamma	Lognormal	Lognormal	Lognormal	Lognormal
	Parameters	$\alpha = 3.55$	$\alpha = 4.06$	$k = 2.51$	$\alpha = 4.13$	$\mu = 0.36$	$\mu = 0.34$	$\mu = 0.33$	$\mu = 0.29$
		$\beta = 2.79$	$\beta = 4.00$	$\lambda = 1.40$	$\beta = 2.64$	$\sigma = 0.60$	$\sigma = 0.63$	$\sigma = 0.66$	$\sigma = 0.68$
Touristic trip distance	Best Fit Distribution	Lognormal	Gamma	Gamma	Gamma	Gamma	Lognormal	Lognormal	Lognormal
	Parameters	$\mu = -0.07$	$\alpha = 3.33$	$\alpha = 3.65$	$\alpha = 3.08$	$\alpha = 2.62$	$\mu = 0.12$	$\mu = 0.21$	$\mu = 0.16$
		$\sigma = 0.63$	$\beta = 3.56$	$\beta = 3.46$	$\beta = 2.25$	$\beta = 1.84$	$\sigma = 0.67$	$\sigma = 0.70$	$\sigma = 0.70$
Commuting trip duration	Best Fit Distribution	Lognormal	None[1]	Gamma	None[1]	Lognormal	Lognormal	Lognormal	Lognormal
	Parameters	$\mu = 2.03$	-	$\alpha = 5.15$	-	$\mu = 2.21$	$\mu = 2.22$	$\mu = 2.21$	$\mu = 2.18$
		$\sigma = 0.56$	-	$\beta = 0.60$	-	$\sigma = 0.59$	$\sigma = 0.62$	$\sigma = 0.63$	$\sigma = 0.65$
Touristic trip duration	Best Fit Distribution	None[1]	None[1]	None[1]	Lognormal	Lognormal	Lognormal	Lognormal	Lognormal
	Parameters	-	-	-	$\mu = 3.03$	$\mu = 2.75$	$\mu = 2.80$	$\mu = 2.77$	$\mu = 2.69$
		-	-	-	$\sigma = 0.83$	$\sigma = 0.79$	$\sigma = 0.90$	$\sigma = 0.77$	$\sigma = 0.75$
System diameter		4.71	2.85	2.33	5.06	8.57	14.28	23.29	11.17

Notes: 1. "None" means that none of the three distributions have K-S statistic lower than 0.05

2. Model parameters corresponds to the parameters in Table 3.

This study further evaluated the goodness-of-fit plots to gain more insights of the difference between the smaller and larger systems. Figure 2.4 compared the goodness-of-fit from different perspectives for one smaller system (Seattle) and one larger system (Chicago). The P-P plots show that these models fit well at the distribution center, while the Q-Q plots reveal a lack of fit at the distribution tail. Because the bike share users of larger systems (e.g., Chicago) can potentially travel longer due to the larger spatial coverage (though only a small percentage of trips are long-distance), the variation at the distribution tail would be more significant and thus may influence the selected distribution models. It can be observed from Figure 2.5 that Chicago has a wider-spread station network than Seattle and thus allows users to travel for longer distance. The reasons why a common distribution model is not found for the small systems would be complex. The system size is a major constrain for the users in these small systems. Because such systems have limited number of stations and small system diameters, the users' movement is also largely restricted in the area where bike share stations are available. Additionally, factors such as the stations' geographic layout, urban road network, and accessibility of bike paths may also affect the bike share users' behaviors.

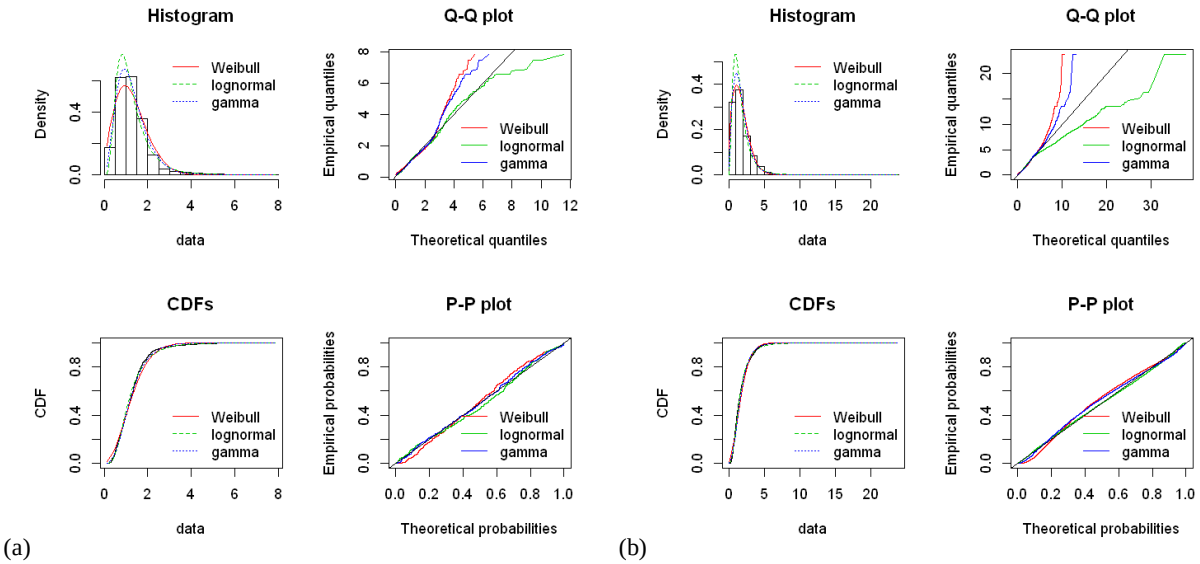
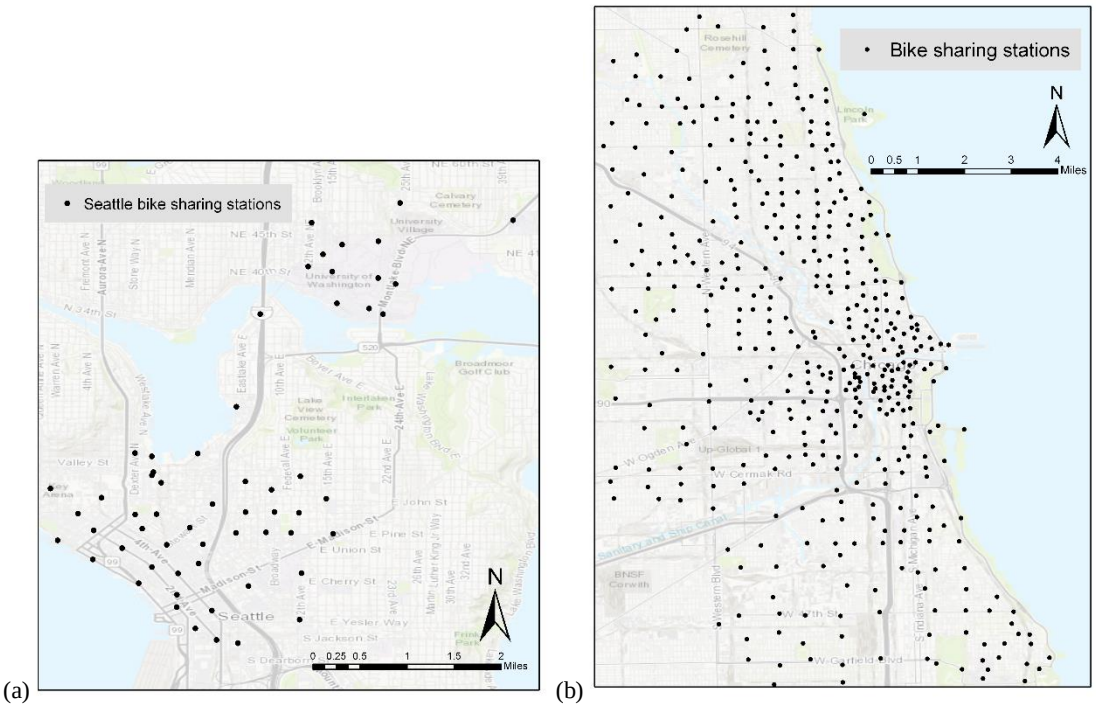


Figure 2.4. Weibull, lognormal, and gamma distribution fitted to commuting trip distance: (a) Seattle, (b) Chicago (graphs for other cities are presented in Figures A. 2 to A. 5 in APPENDIX A)



(a) (b)

Figure 2.5. Station network (black dots) and popular origin-destination pairs (based on historical trip count): (a) Seattle, (b) Chicago

2.3.2 The distribution of trip distance and duration for long trips

As discussed in Section 2.3.1, the major factor that influenced model selection came from the goodness-of-fit in the tail part (long distance/duration). Therefore, analyzing the pattern in the right side tail part is also meaningful. The K-S statistics (Table 2.6) shows that the power law distribution provides a good fit for both the trip distance and duration for the larger systems (Boston, Washington DC, Chicago, and New York). The power law cut-off point (where the power law fitting starts) is determined by optimizing the goodness of fit. Generally, the power law fit for the larger systems starts at longer distance/duration. For those cases that power law exhibits a good fit, the exponent *k* of power law distribution varies from 5.5 to 12.6 for commuting trip distance, 3.4 to 12.1 for tourist trip distance, 5.4 to 16.3 for commuting trip duration, and 2.4 to 3.4 for tourist trip duration. The exponent *k* is basically in the same range in the case of trip distance, for both commuting and touristic trips. This is because that the data only reflects the distance between trip origin and destination. On the contrary, for trip duration, the exponent exhibits significant difference between commuting trips and touristic trips. Touristic trips have smaller exponent in the power law distribution fitting, showing that the probability

that a trip has certain duration drops slower for touristic trips than commuting trips as the trip duration increases. In the study from Jiang et al. (Jiang et al., 2009), they fitted the trip distance of taxi GPS trajectory data to power law. The exponent is around 2.5 for intra-city trajectories and 4-5 for intercity trajectories, respectively, which is similar or smaller than the exponent fitted for bike share trips. This is expected since people is more likely to take long trips by taxi than by bike. Only several smaller bike share systems fit well with the power law distribution, which is due to the restriction from the station network, similar to the analysis in Section 2.3.1. With less restrictions from the bike share system, the mobility pattern in larger bike share systems tends to exhibit power law decay, which is similar to the observation in the research of Jiang et al. (Jiang et al., 2009) . It is also worth noting that the trips in the tail part that are used for the power law fitting only make up a very small portion of the total trips (typically, less than 5%). The proportion is even smaller in the larger systems, which indicates that the probability of long trips is very low. In addition, this study also fitted the right side tail part data to exponential distribution. None of the data fits well to the exponential distribution for any of the cities, neither commuting trips nor touristic trips. The details of the exponential fitting are included in APPENDIX A, Section A.3.

Table 2.6. Results of power law fit at the distribution tail

		Seattle	Los Angeles	Bay Area	Philadelphia	Boston	Washington, DC	Chicago	New York
Commuting trip distance (mile)	Right side cut-off[1]	1.14	0.67	1.26	0.99	1.05	0.86	0.77	0.86
	K-S statistic	0.0518	0.0711	0.1330	**0.0243***	**0.0347***	**0.0287***	**0.0176***	**0.0347***
	Power law cut-off[2]	2.15	2.12	2.10	3.35	5.47	4.07	4.49	6.29
	Exponent k	4.28	12.07	11.70	8.27	12.60	5.51	7.25	11.44
	Percentage of trips in power law fit	4.74%	1.94%	5.81%	2.22%	0.24%	2.41%	1.67%	0.39%
Tourist trip distance (mile)	Right side cut-off	0.81	0.69	0.90	0.69	0.68	0.94	1.04	0.75
	K-S statistic	**0.0496***	0.0573	0.1097	**0.0271***	**0.0407***	**0.0234***	**0.0213***	**0.0230***
	Power law cut-off	1.08	2.09	2.09	3.35	5.37	2.98	9.26	8.10
	Exponent k	3.36	10.63	10.32	8.60	11.18	4.27	8.85	12.09
	Percentage of trips in power law fit	17.77%	0.93%	1.29%	0.72%	0.08%	2.47%	0.03%	0.00%
Commuting trip duration (minute)	Right side cut-off	7.65	0.24	7.45	8.07	5.83	5.58	5.30	5.29
	K-S statistic	**0.0283***	0.1498	**0.0299***	0.1453	**0.0161***	**0.0167***	**0.0132***	**0.0070***
	Power law cut-off	17.16	11.78	19.27	10.00	27.40	23.33	31.47	40.98
	Exponent k	5.36	5.86	16.32	3.74	9.19	6.16	6.94	12.48
	Percentage of trips in power law fit	3.95%	6.42%	0.33%	35.59%	1.17%	3.42%	0.49%	0.16%
Tourist trip duration (minute)	Right side cut-off	11.36	8.67	9.24	12.82	9.45	5.95	9.98	10.40
	K-S statistic	**0.0162***	**0.0258***	0.0515	**0.0242***	**0.0166***	**0.0129***	**0.0050***	**0.0260***
	Power law cut-off	18.32	42.00	10.87	65.00	56.77	119.37	34.48	56.02
	Exponent k	2.46	2.40	2.41	2.55	2.72	3.42	3.07	2.62
	Percentage of trips in power law fit	19.59%	3.28%	12.93%	2.36%	1.37%	1.03%	4.76%	0.97%

1. Right side cut-off is the distance/duration with the highest probability (density) in the probability density plot.

2. Power law cut-off is the lower cut-off for the scaling region estimated by a goodness-of-fit based approach.

3. * Significant at 0.05 level.

4. Model parameters corresponds to the parameters in Table 3.

This study further plotted the Complementary Cumulative Distribution Function (CCDF) graphs to compare between the smaller and larger systems. CCDF graphs are used because each bin in the logarithmic tail part only contains very few samples, resulting in a messy tail in a histogram or probability density function (PDF) plot (Cai et al., 2016; Newman, 2004). Figure 2.6 compares the power law fitting of Seattle (smaller system) and Chicago (larger system). In the log-log range, the straight line of power law fitting (red line) is steeper for Chicago than for Seattle, which is consistent with the result that Chicago has larger exponent *k*.

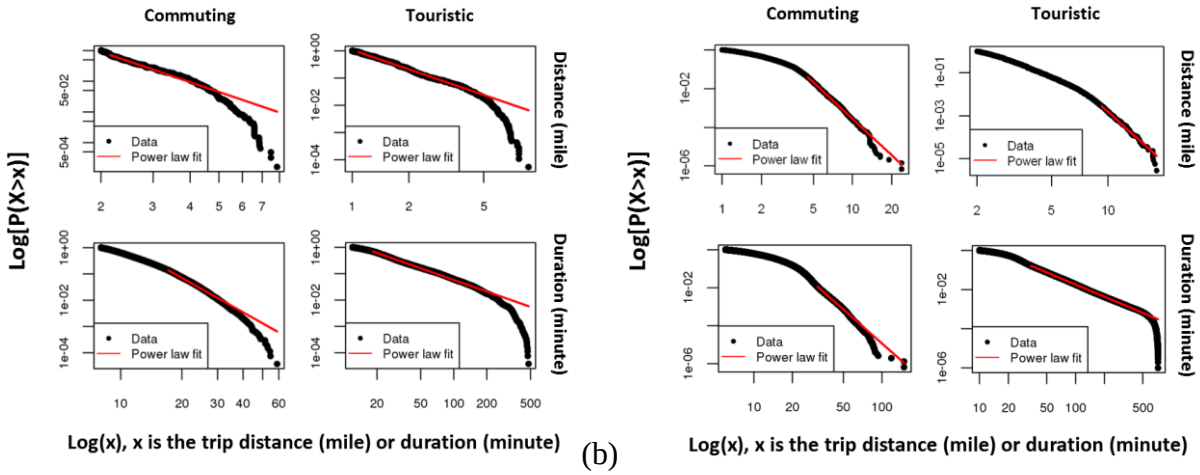


Figure 2.6. CCDF plots of the power law fit: (a) Seattle, (b) Chicago (graphs of other cities are presented in Figures A. 6 to A. 9 in APPENDIX A)

2.3.3 Sensitivity analysis

Because this study used a threshold of 1.3 times of the estimated trip duration as a filter to distinguish commuting and touristic trips, a sensitivity analysis is also conducted to evaluate how this threshold would impact the results. Figure 2.7 provides a comparison between the original data and the filtered data (using different thresholds) on their power law fit. The original data shows a more evident decay in the tail part than the filtered data. Because the bike share trip data does not contain the detailed trace of each trip, the duration brings extra uncertainty to the analysis, for instance, it is difficult to know whether there is idle time during a trip. Even the tail part only makes up less than 6% of the total trips (Table 2.6), it can still contribute to a significant difference when fitting the distribution. It can be observed that the decay and the change of the exponent *k* value of power law are contributed by those "touristic" trips. Table 2.7

and Table 2.8 listed the results of model selection for Seattle and Chicago, with different thresholds. In general, when the filter is closer to 1, the commuting trips will include higher percentage of trips in which the users travel directly from the origin to destination through the shortest path. Seattle's result is more sensitive to this filter than that of Chicago. For Seattle's commuting trip distance, the best model for the whole distance range is gamma distribution when the filter is 1.1 and 1.3 times, while it prefers lognormal distribution when the threshold increases to 1.5 or when there is no filter. The power law fitting to the longer distance trips has a significant K-S statistic only for "no filter" case. Though gamma distribution remains to be a good fit no matter what the threshold is. The shape parameter α and scale (rate) parameter β will decrease slightly (α: 3.84 – 3.06, β: 2.85 – 2.51) as the threshold increases. Regarding Seattle's commuting trip duration, the best fitted model switches from gamma to lognormal when the filter is larger or equal to 1.3. However, none of the right-skewed models have a significant K-S statistic for the non-filter case. For power law fitting, it presents a good fit no matter what the threshold is, but the exponent of the fitted model varies from 8.2 to 2.6. In contrast to Seattle, the threshold did not affect the model selection for Chicago to a very large extent. Lognormal remains to be the best model and power law fits well to the tail in all the cases when the threshold changes. For commuting trip distance, the lognormal model has the mean (log) μ in the range of 0.28 – 0.33 and standard deviation (log) σ in the range of 0.65 – 0.68; while the power law exponent k is in the range of 7.6 – 8.4. The power law exponent k for commuting trip duration may change a lot between nonfiltered and filtered data, since the nonfiltered data has an exponent k of 3.11, while the filtered data has exponent k in range of 6.9 to 8.6. Therefore, when selecting a distribution to model the trip duration of the bike share user's behaviors, those smaller systems (e.g., Seattle) should be treated more cautiously. Again, the variations or sensitivities of the model selection or model parameters are more likely to come from the bike share station networks, which could be the reason why it is difficult to find a common distribution model for those smaller systems.

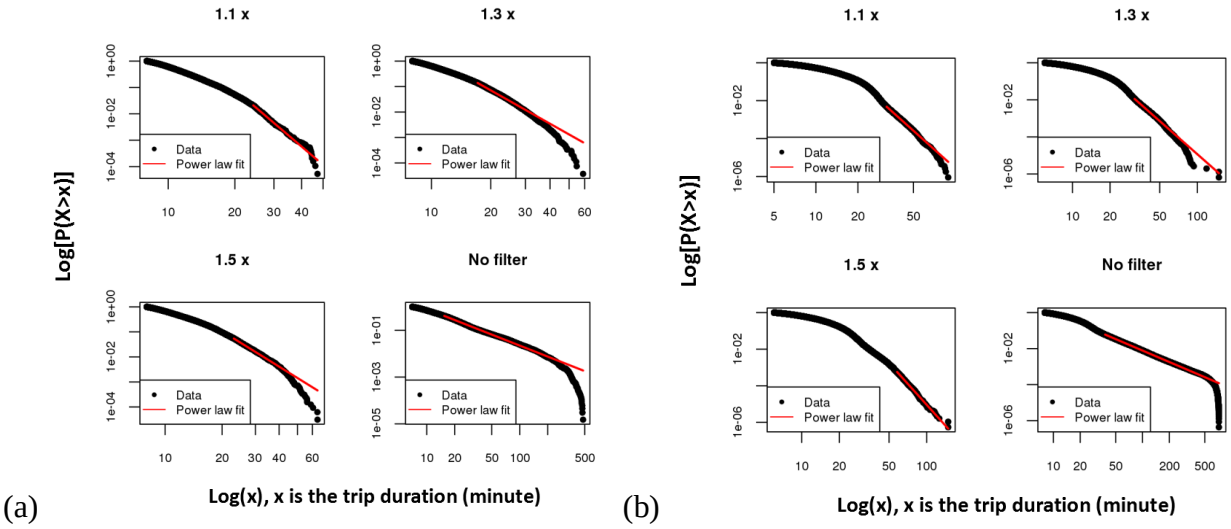


Figure 2.7. CCDF plots of the power law fit to commuting trip duration with different thresholds: (a) Seattle, (b) Chicago

Table 2.7. Results of sensitivity analysis, Seattle

		Filter		1.1	1.3	1.5	No filter
Commuting trip distance (mile)	Weibull	Parameters	Shape	1.8797	1.8512	1.8233	1.6971
			Scale	1.4489	1.4361	1.4234	1.3756
		Goodness-of-fit statistics/criteria	K-S statistic	0.0699	0.0658	0.0638	0.0643
			AIC	81034.01	107316.3	124027.2	186907.1
			BIC	81051.28	107334.1	124045.3	186926
	Gamma	Parameters	Shape	3.6446	3.5464	3.4488	3.0631
			Scale	2.8458	2.7932	2.7399	2.5124
		Goodness-of-fit statistics/criteria	K-S statistic	**0.0420***	**0.0382***	**0.0365***	**0.0377***
			AIC	76656.79	101513.2	117357.5	176672.1
			BIC	76674.07	101531.1	117375.6	176691
	Lognormal	Parameters	meanlog	0.1040	0.0912	0.0782	0.0262
			sdlog	0.5433	0.5500	0.5573	0.5897
		Goodness-of-fit statistics/criteria	K-S statistic	0.0585	0.0504	**0.0468***	**0.0395***
			AIC	75912.96	100235.9	115626.1	172444.9
			BIC	75930.23	100253.8	115644.2	172463.8
	Power Law		Right side cut-off point	1.1615	1.1420	1.1442	0.8195
			K-S statistic	0.0657	0.0518	0.0504	**0.0425***
			Lower cut-off for the scaling region	2.1468	2.1468	2.1468	1.4428
			Exponent k	4.1651	4.2793	4.2616	4.0146

Table 2.7 continued

			No. of points in tail part	3205	4454	5165	26861
Commuting trip duration (minute)	Weibull	Parameters	Shape	1.9528	1.8550	1.7833	0.9983
			Scale	9.4031	10.0013	10.4639	16.7761
		Goodness-of-fit statistics/criteria	K-S statistic	**0.0493***	0.0586	0.0622	0.1253
			AIC	234836.5	320810.9	379173.7	718543.6
			BIC	234853.7	320828.8	379191.8	718562.5
	Gamma	Parameters	Shape	3.7133	3.4615	3.2475	1.2754
			Scale	0.4471	0.3917	0.3508	0.0760
		Goodness-of-fit statistics/criteria	K-S statistic	**0.0214***	**0.0296***	**0.0348***	0.1312
			AIC	231495.9	315744.1	373111.4	715311.8
			BIC	231513.1	315762	373129.5	715330.7
	Lognormal	Parameters	meanlog	1.9762	2.0276	2.0635	2.3803
			sdlog	0.5428	0.5589	0.5752	0.8267
		Goodness-of-fit statistics/criteria	K-S statistic	**0.0393***	**0.0319***	**0.0245***	0.0537
			AIC	231530.2	314757.4	371196	678694.2
			BIC	231547.5	314775.2	371214.1	678713.1
	Power Law		Right side cut-off point	7.3525	7.6474	7.2725	6.1040
			K-S statistic	**0.0213***	**0.0283***	**0.0257***	**0.0214***
			Lower cut-off for the scaling region	24.2198	17.1621	22.8324	15.4299
			Exponent k	8.1714	5.3556	5.6912	2.5627
			No. of points in tail part	396	3715	1835	27158

1. Right side cut-off is the distance/duration with the highest probability (density) in the probability density plot.

2. Power law cut-off is the lower cut-off for the scaling region estimated by a goodness-of-fit based approach.

3. * Significant at 0.05 level.

4. Model parameters corresponds to the parameters in Table 3.

Table 2.8. Results of sensitivity analysis, Chicago

		Filter		1.1	1.3	1.5	No filter
Commuting trip distance (mile)	Weibull	Parameters	Shape	1.6093	1.6144	1.6013	1.5664
			Scale	1.9272	1.9218	1.9067	1.8476
		Goodness-of-fit statistics/criteria	K-S statistic	0.0593	0.0547	0.0528	**0.0484***
			AIC	3913715.2	5717157.5	6783809.6	9323442.1
			BIC	3913739.5	5717182.6	6783835.1	9323468.2
	Gamma	Parameters	Shape	2.5491	2.5535	2.5170	2.4287
			Scale	1.4867	1.4937	1.4828	1.4734
		Goodness-of-fit statistics/criteria	K-S statistic	**0.0497***	**0.0462***	**0.0447***	**0.0395***
			AIC	3820833.9	5585434.3	6628840.4	9108608.9
			BIC	3820858.2	5585459.4	6628865.8	9108635
	Lognormal	Parameters	meanlog	0.3303	0.3278	0.3175	0.2800
			sdlog	0.6551	0.6563	0.6619	0.6767
		Goodness-of-fit statistics/criteria	K-S statistic	**0.0187***	**0.0176***	**0.0184***	**0.0174***
			AIC	3764514.8	5514671.9	6546544.6	9007262.6
			BIC	3764539.1	5514697	6546570	9007288.7
	Power Law		Right side cut-off point	1.0000	0.7733	0.7776	0.7907
			K-S statistic	**0.0241***	**0.0176***	**0.0158***	**0.0182***
			Lower cut-off for the scaling region	4.4925	4.4931	4.3962	8.2686
			Exponent k	7.6295	7.2541	6.8692	8.4114
			No. of points in tail part	43120	58166	76369	3625
Commuting trip duration (minute)	Weibull	Parameters	Shape	1.7504	1.7394	1.7063	1.2125
			Scale	11.6227	12.3974	12.8541	16.4961
		Goodness-of-fit statistics/criteria	K-S statistic	**0.0466***	**0.0439***	**0.0426***	0.0869
			AIC	8862380.8	13275483	16020129	25386248
			BIC	8862405.1	13275508	16020154	25386274
	Gamma	Parameters	Shape	2.8790	2.8419	2.7623	1.8332
			Scale	0.2798	0.2588	0.2423	0.1198
		Goodness-of-fit statistics/criteria	K-S statistic	**0.0339***	**0.0328***	**0.0329***	**0.0458***
			AIC	8785794.8	13166770	15885710	25017980
			BIC	8785819.2	13166795	15885736	25018007
	Lognormal	Parameters	meanlog	2.1477	2.2101	2.2417	2.4307
			sdlog	0.6207	0.6266	0.6358	0.7407

Table 2.8 continued

	Goodness-of-fit statistics/criteria	**K-S statistic**	**0.0236***	**0.0274***	**0.0280***	**0.0216***
		AIC	8769914.6	13152296	15863341	24423729
		BIC	8769938.9	13152321	15863366	24423755
	Power Law	Right side cut-off point	4.9947	5.2965	5.2826	7.0246
		K-S statistic	**0.0105***	**0.0132***	**0.0126***	**0.0047***
		Lower cut-off for the scaling region	32.1333	31.4667	58.0167	38.0667
		Exponent k	7.6032	6.9416	8.5796	3.1133
		No. of points in tail part	5396	17134	1048	139703

1. Right side cut-off is the distance/duration with the highest probability (density) in the probability density plot.

2. Power law cut-off is the lower cut-off for the scaling region estimated by a goodness-of-fit based approach.

3. * Significant at 0.05 level.

4. Model parameters corresponds to the parameters in Table 3.

2.4 Conclusion and Future Work

This study analyzed the mobility patterns of bike share users by identifying the distributions of the trip distance and duration. Because the bike share trip data used in the analysis only contain the time and location information of the trip origins and destinations, this work estimated the trip distance using the Google Maps Distance Matrix API and also separated the data into commuting trips and touristic trips. No common distributions are found that can describe the bike share system for all cities. For the cities with larger bike share systems (i.e., Boston, Washington DC, Chicago, and New York), both the trip distance and the trip duration follow a lognormal distribution. For trip distance, the mean (μ) of the lognormal distribution is between 0.29 and 0.36 miles for commuting trips and between 0.12 and 0.21 miles for touristic trips. For trip duration, the mean (μ) is around 2.2 minutes for commuting trips and 2.8 minutes for touristic trips. For the cities with smaller bike share systems, the best-fit model varies among different cities and different cases. The systems' spatial coverage is a major factor that influences the model selection. Small system size (diameter and number of stations) can restrict bike share users' mobility pattern because the users are required to return the bike to a station. The major difference between the distance/duration distribution of smaller and larger systems lies in the right side tail part. Similar to other transportation modes, trips by shared bikes also exhibit power law decay at the long distance/duration range, although the part of trips following the power law distribution only makes

up a very small portion of the total trips. The exponent k of power law distribution varies from 5.5 to 12.6 for commuting trip distance, 3.4 to 12.1 for tourist trip distance, 5.4 to 16.3 for tourist trip duration, and 2.4 to 3.4 for tourist trip duration. In this study, the trips whose duration is within 1.3 times of the duration estimated by the Google Maps API are defined as the commuting trips and trips taken longer than that as touristic trips. In the sensitivity analysis of this threshold, this study finds that the model selection and model parameters are more sensitive to this threshold for Seattle (a smaller system) than Chicago (a larger system). The distribution fitting is also more sensitive to the threshold for trip duration than for trip distance.

While this study fills the gap of analyzing the travel patterns of bike share trips considering both short and long trips, the limitations of this work should be noted for future research. A major limitation of this analysis is the limited information contained in the raw data (having only trip origin and destination time and location). If the trajectory data of the bike share trips become available, the actual trip distance and duration can be more accurately calculated. In addition, more detailed analysis about the transportation network, pricing scheme, user preference, and other socio-geographic information may also help explain the difference in trip distance and duration distributions among different cities. Lastly, this research focuses on the travel patterns of trips using station-based bike share programs, which may have different travel behaviors compared to the trips using the newly emerged dockless bike share programs (Shaheen et al., 2013a). Future research evaluating the difference between the two types of programs can provide a more complete understanding of the travel patterns of bike share users.

3. QUANTIFYING THE ENVIRONMENTAL BENEFITS OF BIKE SHARE SYSTEMS

The main content of this chapter has been published in a journal paper: Kou, Z., Wang, X., Chiu, S. F. A., & Cai, H. (2020). Quantifying greenhouse gas emissions reduction from bike share systems: a model considering real-world trips and transportation mode choice patterns. *Resources, Conservation and Recycling*, 153, 104534.

3.1 Introduction

Evaluating the benefits of bike share systems is important for the cities to make decisions on supporting the development of bike share programs. Bike share can potentially bring social and environmental benefits such as saving transportation time (faster than walking and even driving in highly congested areas) and expenses, alleviating traffic congestion, reducing greenhouse gas (GHG) emissions and air pollutants, and improving multimodal transport connections (Drynda, 2014; Shaheen et al., 2010). Jäppinen et al. (2013) evaluated the spatial impacts of a hypothetical BSS on public transit travel time and estimated that bike share can reduce public transit travel time by an average of 10% in the Greater Helsinki area in Finland. Faghih-Imani et al. (Faghih-Imani et al., 2017) analyzed bike share trips and taxi trips in New York City and found that, for short trips (less than 3 km), traveling by bike share is either as fast as or faster than traveling by a taxi. Bullock et al. (Bullock et al., 2017) conducted a survey of bike share users in Dublin, Ireland and showed that the BSS contributed to the urban economy because the journey time savings and improved connectivity from bike share increased the productivity of local economic activities. They found that these benefits are more significant than those of modal shifts from cars to bicycles. Using causal inference models, Hamilton and Wichman (Hamilton & Wichman, 2018) proved that the introduction of a bike share system reduced traffic congestion by about 4% in Washington. DC. Fuller et al. (Fuller et al., 2013) analyzed the data from a telephone survey about people's biking behaviors and concluded that there is a higher likelihood of cycling for people who live in areas where bike share is available, compared to people who have privately owned bicycles but have no access to bike share programs. Buehler and Hamre carried out a survey of bike share users and local businesses near the bike share stations in Washington DC, U.S. (Buehler & Hamre, 2015).

They found that 73% of the users chose bike share to save travel time, while 25% of the users were motivated by cost savings.

Out of these benefits, this chapter focuses on evaluating the environmental benefit of BSSs, which is a major benefit of BSSs since bike share trips do not generate tailpipe emissions like car trips. The environmental benefit of bike share comes from the fact that bike share may replace another transportation mode which has worse environmental impacts, e.g., cars. A few studies have tried to use bike share trip data to quantify the benefits of BSSs regarding car travel reduction (Fishman et al., 2014) and carbon emission reduction (Zhang & Mi, 2018). However, these studies estimate the mode replacement based on arbitrary assumptions (Zhang & Mi, 2018) or the assumption that all the bike share trips have the same probability of replacing a certain mode (Fishman et al., 2014) (more details are provided in the Literature Review part of this chapter). These studies did not consider the variations of user behaviors, i.e., people's heterogeneous transportation mode choices under different conditions. For example, a person may prefer walking for a shopping trip but prefer driving for a commuting trip, even though the trip distances are the same. Therefore, detailed trip information such as trip distance and trip purpose needs to be considered for more accurate modeling.

This study proposed a Bike Share Emission Reduction Estimation Model (BS-EREM), which quantifies the greenhouse gas (GHG) emission reductions from BSSs with consideration of the detailed bike share trip information, such as trip distance, trip purpose, and time of the day for the trips. Mode substitution of bike share trips is simulated based on historical travel patterns learned from trip data in National Household Travel Survey (NHTS). The proposed model is then applied to compare the GHG emission reductions of eight bike share systems in the United States, including New York, Chicago, Boston, Philadelphia, Washington D.C., Los Angeles, San Francisco, and Seattle.

The rest of this chapter is structured as follows: Section 3.2 discusses the relevant literature, research gaps, and the contributions made by this study; Section 3.3 provides detailed information about the input data and methodology of the proposed BS-EREM; In Section 3.4, the overall and unit (such as per trip and per bike) GHG emission reduction, as well as the spatial patterns of

different bike share stations' contributions to GHG emission reduction are analyzed by applying the proposed model to real-world trip data; a sensitivity analysis is carried out in Section 3.5 to discuss the impacts of key parameters and assumptions of the BS-EREM; lastly, Section 3.6 summarizes the findings, then discusses the limitations of this work and future research directions.

3.2 Literature review

In order to estimate the replaced modes for bike share trips, researchers have tried to collect modal shift information through user surveys. Martin and Shaheen (Martin & Shaheen, 2014) collected bike share user survey data in Washington, D.C. and Minneapolis. They analyzed the geospatial distribution of the modal shift due to bike share and concluded that higher proportion of bike share trips replaced public transit in urban periphery than in urban areas with high population density. The analysis from Nair et al. (Nair et al., 2013) shows that bike share stations closer to public transit facilities often exhibit higher utilization. Shaheen et al. (Shaheen et al., 2013b) analyzed the modal shift of bike share by conducting user surveys for BSSs in four cities: Montreal, Toronto, Twin Cities, and Washington, D.C. The survey results showed that bike share users shifted from all transportation modes and the modal shifts vary in different cities. About 40% to 50% of the respondents reported that their use of bike share reduces their bus and rail usage in Montreal, Toronto and Washington D.C., while in Twin Cities, 15% of the respondents reported increasing rail usage. In all four cities, bike share was found to reduce car and taxi usage. However, such survey results only provide qualitative information but not quantitative data for the changes (e.g., it shows the trend on whether bike share increases/decreases the use of other modes but does not quantify the changes). Similarly, another study also reveals that bike share reduced the usage of personal automobiles among about 50% of the bike share users in five cities: Twin Cities, Salt Lake City, Montreal, Toronto, and Mexico City (Susan et al., 2014). Two surveys were conducted for the BSS in Washington, DC, separately targeting the subscription members (users who purchase monthly/annual plans) (LDA Consulting, 2016) or casual users (those who only used the service with one-day or five-day plans) (Borecki et al., 2012). The survey results show that 39% of the members and 53% of the casual users who responded to the surveys used bike share to replace walking trips for their last bike share trips.

Few studies quantified the environmental benefits of bike share programs using mode substitution information. The research of Fishman et al. (Fishman et al., 2014) was one of the studies that first analyzed the bike share benefits with consideration of mode substitution. They used the bike share user survey data to estimate the mode substitution. In these user surveys which were carried out by the BSS operators, the BSS users were asked about their alternative travel modes, if not using bike share, to calculate the aggregated substitution rate (percentage) of different transportation modes. For instance, if 40% of the survey participants chose vehicles as their alternative mode, it is assumed that 40% of the bike share trips substituted vehicle trips. Based on the substitution rates, the authors estimated and compared the car travel distance reduction by BSSs in Melbourne, Brisbane, Washington, D.C., Minnesota, and London. The substituted car travel at each BSS is calculated as the car substitution rate multiplied by the total annual trip distance, which is estimated using the bike share trip duration and an assumed biking speed of 12 km/hour. However, this approach assumes that 1) the transportation mode substitution is independent of the trip distance and trip purpose, 2) the percentage of trip substitution is equivalent to the percentage of travel distance substitution, and 3) the trip distance is linearly correlated to the trip duration. These assumptions could lead to biased results in estimating the environmental benefits of bike share programs. First, trip distance and purpose do affect people's transportation mode selection. For instance, for long-distance trips, people are more likely to prefer car or public transit to walking. Additionally, the transportation mode choice would be different if the trip is for sightseeing instead of commuting (USDOT, 2009). The user survey data doesn't include information for trip distance and purpose. Therefore, assuming that all the bike share trips follow the same mode substitution rates could lead to inaccurate results. Second, because the travel distance from trip to trip could vary significantly, the percentage of trip substitution (information asked in the survey) is not equivalent to the percentage of travel distance substitution (information the calculation is based on). Third, while the trip distance is linearly related to the trip duration for some trips (e.g., when the riders go direct from point A to point B, such as commuting), the distance of trips with detours or multiple stops (e.g., leisure trips) cannot be accurately estimated using the trip duration (Kou & Cai, 2019). Therefore, to better understand the transportation emission reduction contributed by bike share programs, the analysis needs to consider transportation mode substitution integratedly with trip distance and trip purpose. In another study which estimated the environmental benefits of the BSSs in Shanghai, China, Zhang and Mi (Zhang & Mi, 2018) set a distance threshold of

1km to divide the bike share trips into two groups: trips that are shorter than 1 km, which were assumed as not replacing car travel and therefore generating zero environmental benefits, and trips that are longer than 1 km, which were assumed as replacing car trips and providing environmental benefits. The environmental benefits from these trips were calculated by multiplying the total trip distance with impact factors (e.g., energy saving per km or emission reduction per km). Although this study considered trip distance in the analysis, the 1 km distance threshold is set arbitrarily. Trips that are less than 1km could also replace a car trip while trips longer than 1 km could also substitute public transit trips or even walking trips. Anderson (Anderson, 2015) assessed a proposed bike share program in Portland, Maine in the United States, and concluded that the potential bike share program can improve air quality and rates of physical activity for bike share users. The analysis is based on several assumed parameters such as annual trips, average trip distance, and average minutes of physical activity per trip. The estimated benefits still need to be furthered verified with real-world data.

In summary, existing studies quantifying the environmental benefits of BSS have two major limitations. First, the calculation of emission reduction from bike share trips is based on simplified assumptions (for example, "all bike share trips longer than 1 km will replace car trips (Zhang & Mi, 2018)"). Second, existing studies often replace the concept of "percent of miles" with "percent of users", when mode substitution data from user surveys are used in the calculation. While user surveys can provide useful information about mode substitution, existing user surveys only reported the share of respondents claiming that they use bike share to replace a certain mode. However, not all users take the same number of bike share trips and not all trips have the same length. It could lead to biased results if directly applying such "percent of users" information as if they were "percent of miles" to quantify the environmental benefits. A user who take more trips and longer trips should be given a higher weight. Ideally, the historical bike share trips made by each survey participant should be used to link the detailed trip information with mode substitution. However, because the trip data and the user survey data are collected separately and anonymously, such linked data is not available at a large scale. Additionally, because user surveys are self-reported data, the accuracy of the data is always a concern (Cohen & Shaheen, 2018). Therefore, better modeling approaches are needed to evaluate the environmental benefits of BSSs using real-

world data and holistically evaluate the transportation mode substitution by bike share trips, considering detailed trip information such as trip distance, purpose, and start time.

To address the above discussed gaps, this study proposes a Bike Share Emission Reduction Estimation Model (BS-EREM) to quantify the greenhouse gas emission reductions by bike share trips through probabilistically simulating bike share trips' mode substitution based on trip distance, trip purpose, time of the day for the trip, and historical travel patterns (mode choice distributions) before launching bike share programs. The model is applied to compare the life cycle greenhouse gas emission reduction from bike share programs in eight U.S. cities and also evaluated the spatial patterns of reduced GHG emissions at the station level in different cities. Compared to the existing literature, the unique contributions made by this study are: 1) this study proposed a model to estimate the mode substitution of bike share trips, with the consideration of trip distance, trip purpose, trip start time, and the public transit accessibility near bike share stations; 2) the mode substitution estimation is based on travel survey data and real-world bike share trip data, thus, the proposed method can be generally applied to all cities that have such data available; and 3) the analysis not only quantifies the overall environmental benefits at the system level, but also evaluated the environmental benefits for unit distance travelled and the spatial distribution of the environmental benefits at the bike share station level, which provides insights for BSS station siting and planning. Results from this study can inform decision makers, city planners, and BSS operators to better develop, deploy, and operate BSS programs to improve transportation sustainability.

3.3 Data and Methods

3.3.1 Input data

The proposed BS-EREM model integrates data from multiple sources to estimate the environmental benefits of BSSs. The input data includes (1) bike share trip data (Section 3.3.1.1), (2) National Household Travel Survey (NHTS) data (Section 3.3.1.2), which provides historical transportation mode choice information prior to launching BSS programs, (3) public transit stations/stops near bike share stations (checking whether it is possible for a bike share trip to substitute a public transit trip, Section 3.3.1.3), and (4) GHG emission factors used for the

calculation of trip environmental benefits given different mode substitution by BSS trips (Section 3.3.1.4). This section explains each type of data while the BS-EREM model is presented in Section 3.3.2.

3.3.1.1 Bike share trips in eight U.S. cities

Same as Chapter 2, this study analyzed the bike share trip data from programs located in eight cities: Seattle (Pronto Cycle Share), Los Angeles (Metro Bike Share), Bay Area (Ford Gobike), Philadelphia (Indego Bike Share), Boston (Hubway, now rebranded as Blue Bikes), Washington D.C. (Capital Bike Share), Chicago (Divvy), and New York City (Citi Bike Share). To compare all eight BSSs, this study used the trip data in year 2016, because Seattle's Pronto Cycle Share was closed after March, 2017. All eight systems are station-based. These bike share programs have made their trip data publically available, which include the timestamps and locations of the origin and destination of each trip. Because no trip distance nor detailed bike trajectory data is provided, this study estimated the trip distance between each pair of origin and destination stations using Google Maps Distance Matrix API, which could more accurately estimate the distance travelled by bike on urban street networks. Many of previous studies used trip displacement (e.g., the Great Circle Distance) to measure human movements (González et al., 2008; Rhee et al., 2011). However, in an urban transportation system, the movement of bikes is restricted by the street networks. The distances estimated by the API can better reflect the bike share trip along the pathways suitable for biking. Compared with driving, bike travels are less affected by congestion (Alter, 2008; Lobo, 2013). Therefore, this study assumed that the travel distance and duration between the same pair of start and end stations stay constant in spite of the trip start time. Round trips, which have the same origin and destination station, are excluded in this analysis due to the difficulty in estimating trip distance. Such round trips only make up 2% to 10% of the total trips in the eight systems. For the overall emission reduction estimation, this study scaled up the results based on the total count of trips, assuming that these round trips share the same pattern as the one-way trips. Because this is an arbitrary assumption, this study has also reported the emission reduction values excluding the round trips.

3.3.1.2 National Household Travel Survey (NHTS)

The National Household Travel Survey (NHTS) (USDOT, 2009) is a nation-wide survey conducted by the U.S. Federal Highway Administration, which records the personal and household travel behaviors of local residents in the United States. Although more recent NHTS data has become available, this study used the 2009 NHTS data to generate the mode substitution rate distributions, because it could better reflect people's transportation mode choice when bike share service was not available (most of the BSSs in the United States were established after the year 2010). There are near 1.05 million trips recorded in NHTS 2009 data, which contains the detailed trip information and the characteristics of the people who are travelling. This study extracted the attributes such as transportation modes, trip purpose, travel time, and travelled distance (in miles) of those trips taken in urban areas (typically with a population density of more than 1,000 persons per square mile (USDOT, 2009)) for each of the eight cities. With such information, the distributions of people's transportation mode choices can be developed for a specific trip distance, travel time, and purpose. This study only used NHTS records in the urban area because the bike share stations analyzed in this study are all located in urban areas defined in the 2016 Topologically Integrated Geographic Encoding and Referencing system (TIGER) data from the U.S. Census Bureau (Thompson et al., 2016), which is also the data source that NHTS used to define urban areas (USDOT, 2009). In the case that a bike share system/station is located in rural area, the mode distribution should be generated using the NHTS trips from rural areas. Figure 3.1 shows the mode choice distributions from the NHTS trip data for all time periods of the day with varied trip distance for commuting and leisure trips in each city. These distributions show that people are more likely to choose walking for short-distance trips (e.g., less than 2 miles) than for long-distance trips (e.g., 10 miles). For the same distance range, less people will choose public transit for leisure trips than for commuting trips. Figure 3.2 depicts the share of different transportation modes in different periods of the day in Los Angeles. The 24 hours of a day are separated into six time intervals. This study group 21:00 to 5:59 as one period because the number of trips during this period is relatively low. For other time of the day, a three-hour interval is used. The mode share varies within a day, especially in the morning and evening. For example, during the period between 6 am and 8:59 am, higher proportion of trips are taken by walking and public transit than other periods such as from 9 am to 5:59 pm. These observations demonstrate the need to include trip distance, purpose, and time of the day in the analysis of trip mode substitution. Therefore, for each city, a total of 12 distributions are developed (2 trip purposes $\times$ 6 time intervals).

3.3.1.3 Public transit accessibility

Another factor that could affect people's mode choice is the accessibility of public transit. If there's no public transit facilities (e.g., bus stops, subway stations) near the origin and destination stations of a trip, it is less possible that the bike share trip replaced public transit. Using Google Maps Places API, this study counts the number of public transit facilities within a 200-meter (0.12 miles, about 3 to 4 minutes' walk) buffer zone around each bike share station in each of the eight cities. The impact of choosing 200 meters as the public transit accessible zone is tested in the sensitivity analysis (Section 3.5). The details on how the data is used are described in Section 3.3.2.

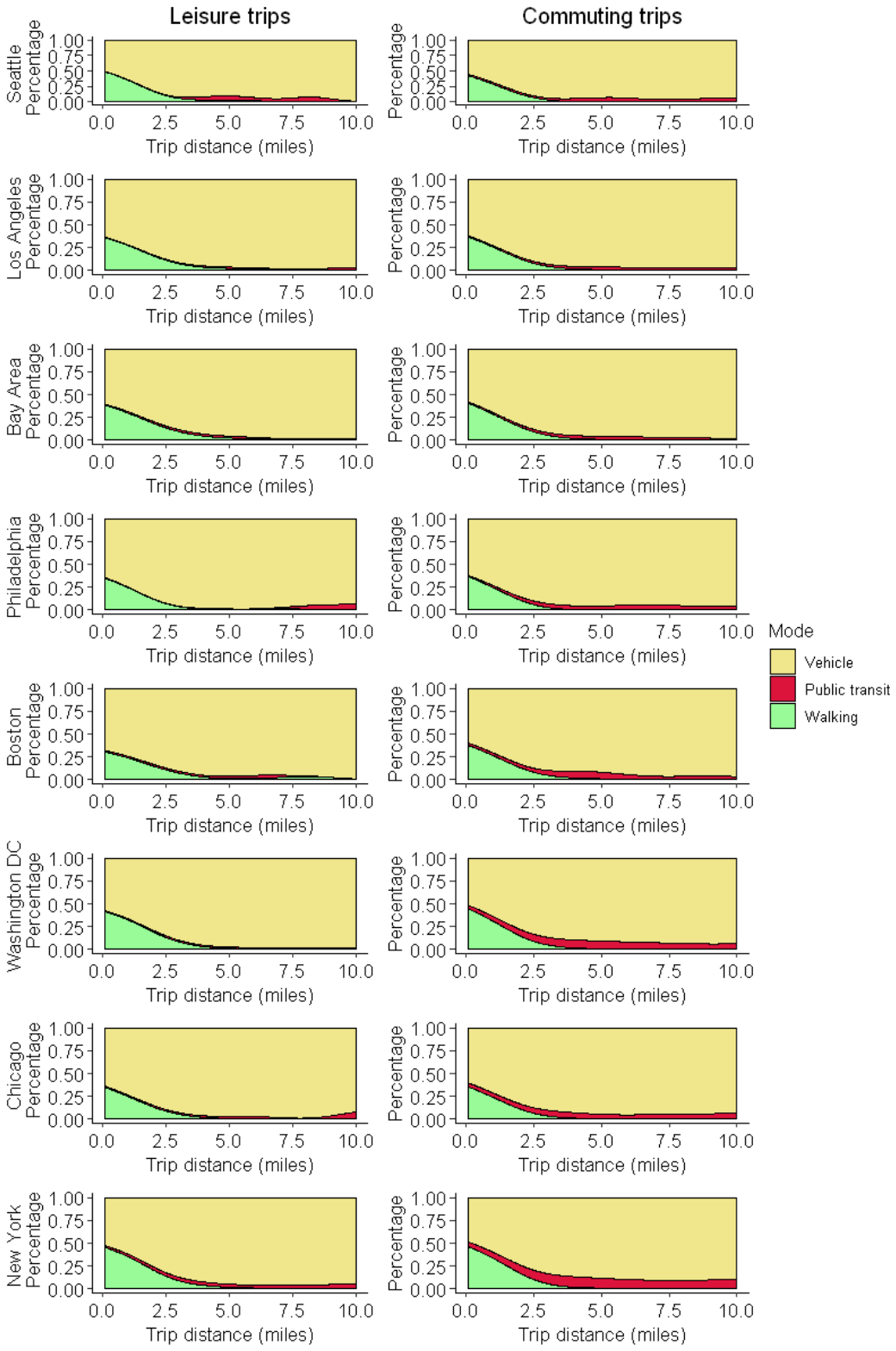


Figure 3.1. Distributions of transportation mode choice regards to different trip distance for commuting trips and leisure trips

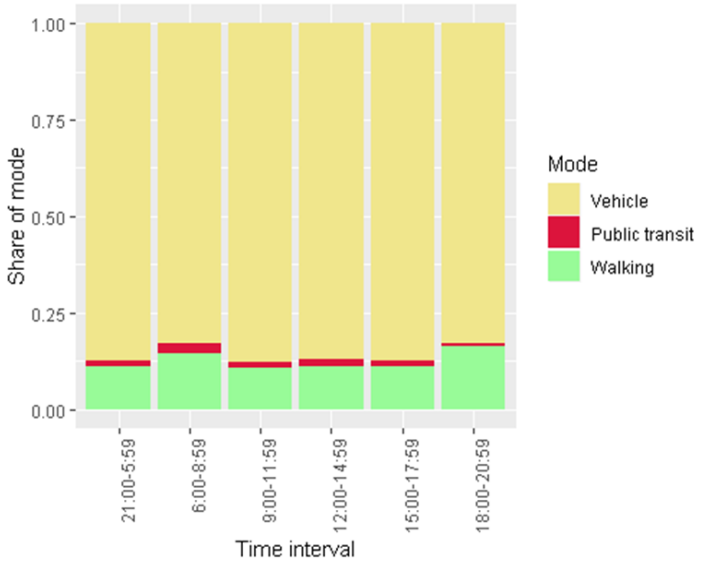


Figure 3.2. Share of transportation mode choice in different time of the day in Los Angeles

3.3.1.4 GHG emission factors

This study use life cycle GHG emission factors from Shreya Dave (Dave, 2010) to calculate the emission reduction associated with each bike share trip (note that, in the life cycle analysis, emissions from the upstream processes besides the trips, such as road constructions, are also allocated in the emission factor calculation). The emission factors from this study are adopted because this study evaluated the emission factors of different transportation modes based on the same analysis scope and assumptions, making the emission factors more comparable to each other than factors drawn from different studies which may be generated based on different assumptions. Because the emission factors developed in this study include multiple types of public transit (e.g., bus and subway) and vehicles (e.g., car, SUV, and pickup), this work takes the weighted average of the emission factors in each transportation mode category to simplify the analysis to include only four transportation modes: walking, bicycling, vehicle, and public transit (Table 3.1). The weighting factors are the corresponding miles travelled using each mode in the NHTS 2017 data (here the 2017 data is used because it can better reflect the distance travelled by different modes in year 2016; NHTS 2017 and 2009 are the two most recent NHTS survey datasets available). Because it is difficult to find more fine-grained emission factor data for all the analyzed cities, this

work use the weighted average emission factors described above as the emission factors for the analysis of all eight cities.

The difference between the life cycle emission factors of the substituted transportation mode and biking is then used to calculate the GHG emission reduction. For example, if a bike share trip substituted a vehicle trip, the emission reduction for this trip would be 408 g CO_2-eq/mile (biking generates 408 g CO_2-eq less than vehicles per passenger mile travelled) multiplied by the trip distance (in miles). How to determine the substituted transportation mode will be discussed in Section 3.3.2.

Table 3.1. Emission Factors used for the GHG emission reduction calculation

Mode	**Sub-category**	**Emission factor (g CO_2-eq/passenger mile travelled)** (Dave, 2010)		**Miles travelled in 2017 NHTS** (U.S.DOT, 2017)	**Weighted average of the emission factor**
Walking	Walking	33			33
Bicycling	Bicycling	33			33
Vehicle	Sedan	382		3,843,053	441
	SUV	446		2,278,726	
	Pickup	619		1,230,229	
Public transit	Bus Average	326		170,342	299
	BART[1]	136	Average of three light rail/subway: 177.7	39,207	
	MUNI[1]	173			
	Green Line[1]	224			

Note: 1 BART: Bay Area Rapid Transit System; MUNI: San Francisco Municipal Railway; Green Line: subway system in Boston; since the emission factors are not available for all the light rail/subway systems in the eight cities, this study used the average emission factors of BART, MUNI and Green Line to represent the unit emission of light rail/subway.

3.3.2 Bike Share Emission Reduction Estimation Model (BS-EREM)

Using the above mentioned data as inputs, this study proposes a Bike Share Emission Reduction Estimation Model (BS-EREM) to estimate the GHG emissions reduction from a bike share trip (Figure 3.3). The BS-EREM model includes two major components: bike share trip purpose estimation and bike share trip mode substitution simulation.

From the NHTS data, the historical trips are first separated into six time groups as specified in Section 3.3.1.2 based on the trip start time. Then for each time group, this work developed the historical mode choice distributions for commuting and leisure trips separately. To differentiate between commuting and leisure trips, the trip speed is compared with the average speed of all trips in this city as a way to infer trip purpose. Trips whose speeds are lower than the average speed are considered as leisure trips (e.g., stop at different locations, take detours to visit different places-of-interest, and not in a hurry to get from the origins to the destinations), while the others are considered as commuting trips (e.g., directly going from the origins to the destinations). The impact of using the average speed as the threshold for trip purpose estimation is tested in the sensitivity analysis (Section 3.5). Given the distance of each bike share trip, the estimated trip purpose, time period of the day, and the corresponding historical distribution of mode choice regarding the specific trip distance, purpose, and time period, the probability of the bike share trip replacing vehicles, walking, or public transit can be obtained to determine the replaced mode in the simulation process. Using the mode choice distributions from NHTS data to simulate the replaced mode by bike share assumes that an average BSS user is the same as an average person in the city. It is possible that BSS users are a self-selected group which has different travel patterns (e.g., having lower percentage of vehicle trips than an average person). The influence of this assumption is tested in the sensitivity analysis (Section 3.5). If there's no public transit facility available near the location of the bike share trip's origin or destination, it is unlikely that this bike share trip replaces public transit. In this case, the mode share of public transit will be set to zero. The pseudo code describing the detailed algorithm to determine the replaced mode in the simulation is provided in the APPENDIX B, Section B.1. The GHG emission reduction from the trip can be calculated using the emission factors and trip distance. This study used the trips in the entire year of 2016 for the simulation and compared the results among different cities.

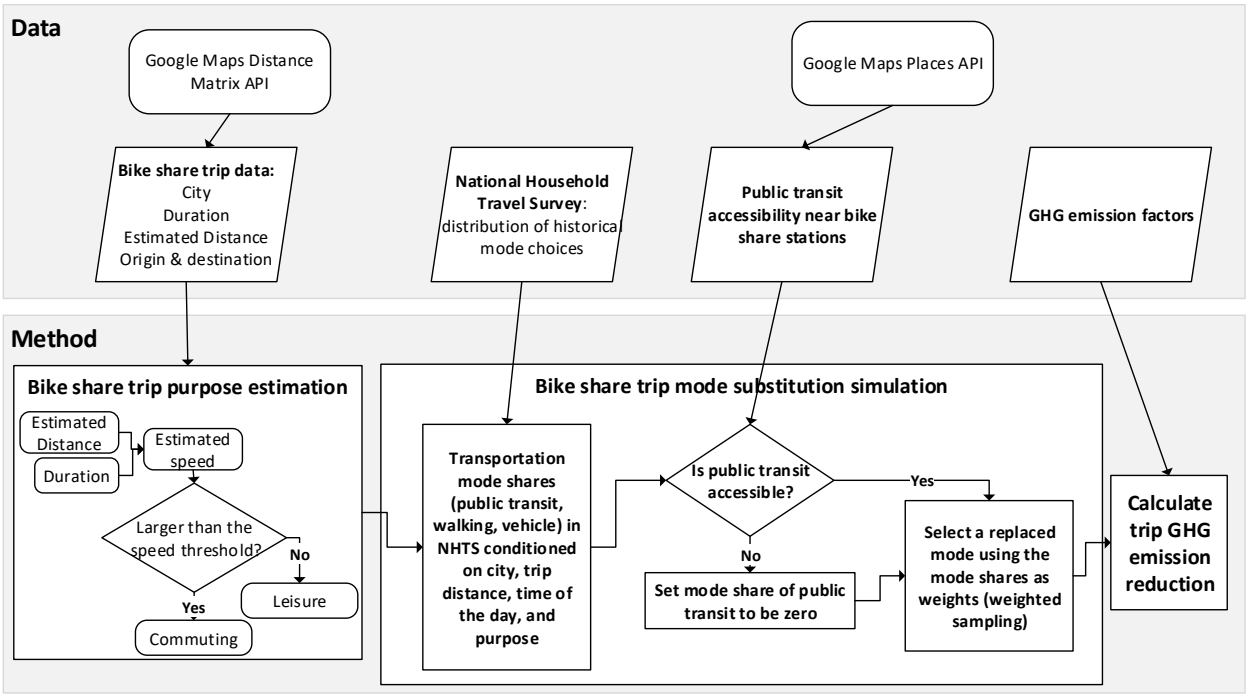


Figure 3.3. Overview of the BS-EREM

3.3.3 Evaluating the emission reduction from BSSs in the context of the transportation sector emissions

After obtaining the total GHG emission reduction using BS-EREM, the results are compared with the total GHG emissions from the transportation sector in these cities. This reflects the BSSs' relative GHG emission reduction contribution to its local transportation systems. The data of GHG emission from the transportation sector is collected from the most recent greenhouse gas inventory report of these cities. The cities do not release such data every year but the data from adjacent years can reflect the general level of transportation sector GHG emissions in these cities (Table 3.2). The transportation sector's GHG emissions have been staying relatively stable based on historical data (for instance, the transportation sector's GHG emissions for City of Chicago in the year 2005, 2010, 2015 are 8.20, 7.95, and 8.05 million MT, respectively (Report, 2017). The changes for 2010 and 2015 compared to 2005 are only -3% and -2%, respectively.

Table 3.2. GHG emissions from the transportation sector in the eight evaluated city

City	GHG emissions from transportation sector (million MT CO_2-eq/year)	Year of the data	Data Source
Seattle	3.58	2014	2014 Seattle Community Greenhouse Gas Emissions Inventory (Kevin, 2014)
Los Angeles	33.74	2010	2015 Environmental Report Card for Los Angeles (Gold et al., 2015)
San Francisco	1.98	2016	2016 San Francisco Geographic Greenhouse Gas Emissions Inventory at a Glance (Program, 2015)
Philadelphia	3.96	2012	Philadelphia Citywide Greenhouse Gas Inventory, 2012 (Spring, 2015)
Boston	1.64	2013	City of Boston Community Greenhouse Gas Inventory 2005-2013 (Boston, 2014)
Washington DC	1.74	2013	District of Columbia Greenhouse Gas Inventory Update 2012-2013 (Department of Energy and Environment, 2012)
Chicago	8.05	2015	City of Chicago Greenhouse Gas Inventory Report, 2015 (Report, 2017)
New York	15.48	2015	Inventory of New York City Greenhouse Gas Emission in 2015 (Pasion et al., 2017)

3.4 Results and Discussions

To understand the GHG emission reduction contributions of BSSs in different cities, this section analyzes the modeling results from two perspectives: the total GHG emission reduction from a BSS in each city and the unit emission reduction by trip, dock, bike, and mile travelled (Section 3.4.1), as well as the spatial distribution of the reduced GHG emissions (Section 3.4.2).

3.4.1 Overall environmental benefits of the BSSs

Table 3.3 listed the basic statistics and modeling results of the bike share trips for each city. The BSSs in New York, Chicago, Boston, and Washington, DC are the four largest station-based systems in the United States (NACTO, 2017). In year 2016, these systems generally have more than 150 stations and generated more than one million bike share trips. These larger systems also have a greater system diameter (indicating spatial coverage), which is defined as the longest Euclidian distance between any two of the bike share stations in a BSS. The total GHG emission reduction in 2016 in these four larger systems are much higher than those in the other four smaller systems (Figure 3.4(a)). New York has the largest amount of total GHG emission reduction: with 10,262,649 bike share trips taken in 2016, New York's BSS contributed to 5,417 tons of GHG emission reduction (in CO_2-eq). In contrast, Seattle's BSS only reduced 41 tons of GHG emissions.

One thing to note is that, this study estimates the emission reduction of round trips based on the assumption that these trips have the same emission reduction per trip. Take New York as an instance, without the consideration of those round trips, the total GHG emission reduction is 5,305 tons CO_2-eq. Although the round trips only take up a very small portion (2% in New York), this part of estimation could be improved if the detailed trajectory data is available.

The emission reduction per trip ranges from 280 to 590 g CO_2-eq (Table 3.3). The total GHG emission reduction is found to correlate linearly to the ridership (total count of trips) in each city with an R squared of 0.998 (Figure 3.4(b)). The linear regression also shows that the marginal emission reduction of one additional trip is 533 g CO_2-eq. Similar linear relationship is also found between the total GHG emission and the total number of bikes (R squared: 0.970, Figure 3.4c). Each bike can contribute to an average of 0.07 to 0.49 tons of GHG emission reduction depending on the system (Table 3.3), while the marginal emission reduction of adding one more bike into the system is 0.51 tons CO_2-eq in a year. In the same way, a linear relationship can be observed between the total GHG emission and the total number of docks (R squared: 0.984, Figure 3.4(d)). The marginal emission reduction of adding one more dock into the system is 0.28 tons CO_2-eq in a year. Also note that in the fitted linear models, the marginal emission reduction (gradient) is mainly influenced by the four larger systems. For the smaller systems such as Seattle and Los Angeles which have less bikes and docks, they also exhibit much lower per-bike (less than 0.1 ton CO_2-eq per bike) and per-dock (less than 0.1 ton CO_2-eq per dock) emission reduction than those larger systems (more than 0.2 ton CO_2-eq per bike and 0.25 ton CO_2-eq per dock). The bikes and docks are utilized less efficiently in these smaller systems. Each bike/dock in these smaller systems on average served less trips compared with that in larger systems, as shown in Table 3. For example, each bike and dock in Seattle on average served 99 trips and 222 trips in 2016, respectively; while the trips per bike and trips per dock could be as high as 503 trips and 979 trips, respectively in New York. The reasons for this need to be further investigated in future research. Regarding trip purpose, commuting trips take up about half of the total trips in the eight cities, ranging from 47% to 54% (Table 3.3). However, the GHG emissions reduction from commuting trips takes up a higher share (54% to 61%) than that from leisure trips, due to combined reasons of travel distance and mode share distribution. Take Los Angeles as an example, the average distance of commuting trips is 1.08 miles, while that of leisure trips is only 0.90 miles. On the other hand, higher

proportion (64.3%) of commuting bike share trips are car-replacing trips than that of leisure trips (60.3% car-replacing trips).

Table 3.3. Basic statistics and analysis results of the bike share systems

City	Seattle	Los Angeles	Bay Area	Philadelphia	Boston	Washington D.C.	Chicago	New York
Total trips in 2016	102,606	184,345	193,506	499,306	1,236,199	2,562,718	3,595,383	10,262,649
Total stations	59	64	74	119	327	407	581	687
Total docks[a]	1,038	1,352	1,357	2,280	5,729	6,720	9,987	20,390
Docks per station[a]	17.59	21.12	18.34	19.16	17.52	16.51	17.19	29.68
Percentage of stations with public transit access	57.63%	67.19%	93.24%	100.00%	96.94%	71.01%	98.28%	57.79%
Percentage of trips used for simulation[b]	91.64%	89.40%	97.58%	91.90%	96.87%	95.36%	95.73%	97.94%
NHTS trips used for simulation[c]	1,619	49,427	29,284	3,350	3,293	14,199	6,096	40,631
Total number of bikes[d]	463	763	421	1,023	1,797	4,305	5,746	10,486
System diameter (miles)[e]	4.71	2.85	2.33	5.06	8.57	14.28	23.29	11.17
Average number of trips per station	1,739	2,880	2,615	4,196	3,780	6,297	6,188	14,938
Average number of trips per dock	99	136	143	219	216	381	360	503
Average number of trips per bike	222	242	460	488	688	595	626	979
Median of trip distance (miles)	1.16	1.2	1.39	1.46	1.43	0.91	1.39	1.31
Average trip distance (miles)	1.27	1.23	1.56	1.7	1.72	1.02	1.71	1.68
Average speed of bike share trips (miles/hour)	6.87	6.24	7.97	7.28	8.21	7.55	7.71	7.92
Percentage of commuting trips	51.92%	48.45%	52.30%	54.37%	52.57%	52.85%	52.88%	46.95%
Total GHG emission reduction (without round trips, ton CO_2-eq)	37.21	46.56	67.43	234.83	668.11	1,275.88	2,000.89	5,305.13
Total GHG emission reduction (including round trips, ton CO_2-eq)[f]	40.60	52.08	69.10	255.51	689.66	1,338.00	2,090.05	5,416.68
Percentage of GHG emission reduction from commuting trips (without round trips)	57.70%	53.93%	60.89%	60.50%	60.14%	61.49%	60.93%	57.16%
Average emission reduction (g CO_2-eq) per mile travelled	324.57	286.59	298.75	340.89	345.02	329.69	352.66	329.48
Average emission reduction (g CO_2-eq) per trip	395.73	282.50	357.12	511.73	557.89	522.10	581.32	527.81

Table 3.3 continued

Average emission reduction (ton CO_2-eq) per station	0.69	0.81	0.93	2.15	2.11	3.29	3.60	7.88
Average emission reduction (ton CO_2-eq) per dock	0.04	0.04	0.05	0.11	0.12	0.20	0.21	0.27
Average GHG emission reduction (ton CO_2-eq) per bike	0.09	0.07	0.17	0.26	0.35	0.28	0.37	0.49

Notes:

a. For Seattle, Bay Area, Boston, and Chicago, the station capacity (number of docks of each station) information is extracted from the station information data (year 2016) released by the system operator; while such data is not available for other four cities: for Los Angeles, Philadelphia, and New York, the station capacity information is obtained from a station status snapshot (real-time station status) in Oct. 15, 2017 from the operators' website; for Washington, DC, the station capacity information is extracted from the bike share station information (Aug. 20, 2018) in Open Data of Washington, DC government[2]. Although the information for the latter four cities mentioned is not exact the station information for year 2016, it can reflect the general level of station capacity of these systems.

b. Round trips are excluded in the simulation

c. Only NHTS survey records in the urban area of the corresponding city were used in the analysis

d. All bikes that showed up in the trip data in year 2016

e. The system diameter is the longest Euclidian distance between any two bike share stations of a BSS at the end of year 2016. Data is collected from (Kou & Cai, 2019)

f. Emission reduction for those round trips is calculated by number of trips multiplied by emission reduction per trip of the those one-way trips

[2] Capital Bike Share Locations, Open Data of Washington, DC, http://opendata.dc.gov/datasets/capital-bike-share-locations

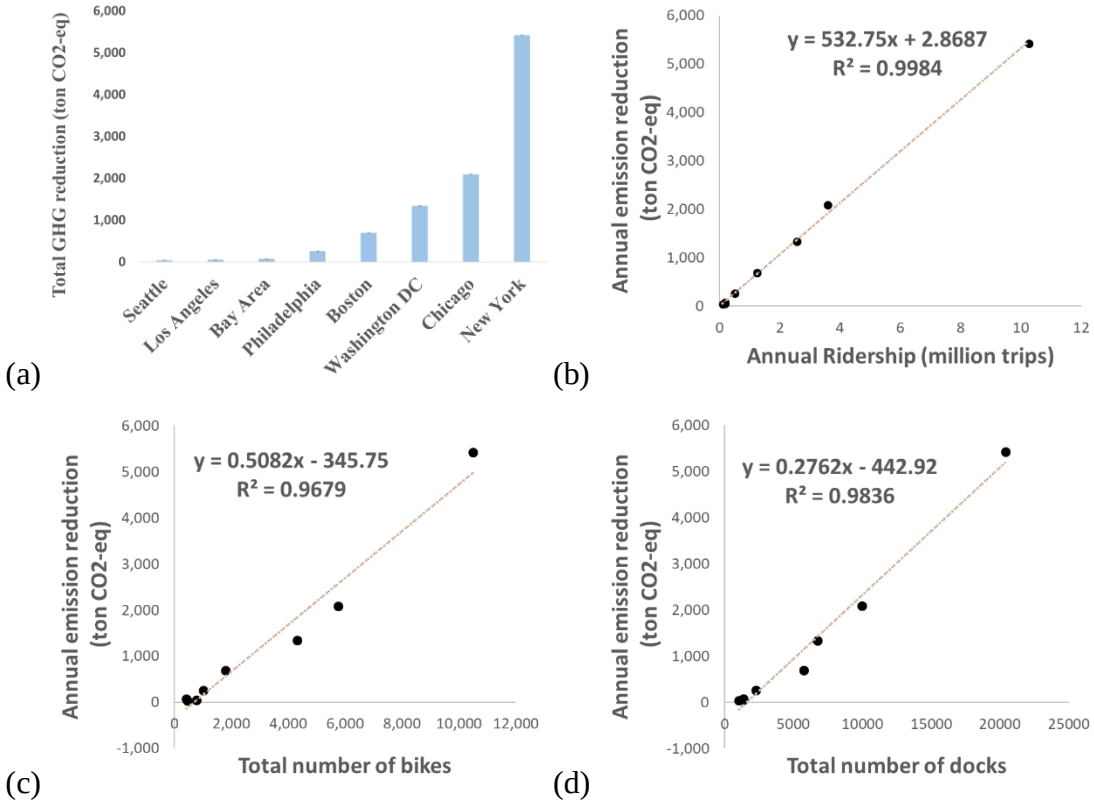


Figure 3.4. Total GHG emission reduction in 2016 of the eight systems (a) and its relationship with (b) annual ridership (count of trips), (c) total number of bikes, and (d) total number of docks

In order to understand the relative environmental benefits of BSSs contributed to these cities, this study also compared the emission reduction from BSS to the total emission from the transportation sector in each city (Figure 3.5(a)). Generally, the emission reduction from BSSs only makes up a small part of the total GHG emission from the transportation sector, which ranges from 0.0002 % to 0.007 % for the four smaller systems and 0.026 % to 0.077 % for the four larger systems. Among these cities, Los Angeles has the lowest ratio because it has a high transportation sector emission but a relatively small BSS.

The GHG emission reduction per mile travelled varies among different cities (Figure 3.5(b)): Chicago, Boston, and Philadelphia have the best performance, which reduced 352.7 g, 345.0 g and 340.9 g CO_2-eq for each mile of bike share trip travelled, respectively. Figure 3.5(c) shows that, in these three cities, higher percentage of bike share trips replaced car trips (74% for Chicago, 75% for Boston, and 75% for Philadelphia), which is the main reason that they have better GHG

emission reduction per mile travelled. However, the GHG emission reduction per bike share trip in Chicago is slightly higher because Chicago's larger bike share station network allows the users to travel longer distance. Generally, the majority of the GHG emission reduction are contributed by relatively short trips (less than 5 miles, Figure 3.5(d)). As shown in Table 3.3, the average and median of bike share trip distance are both less than 2 miles for all the cities. However, in larger systems which have a larger spatial coverage of bike share stations, higher proportion of emission reduction was from trips within 2-5 miles instead of 0-2 miles, compared to those smaller systems. Therefore, the bike share station network plays an important role in affecting the users' travel pattern and the BSSs' ability to reduce GHG emissions.

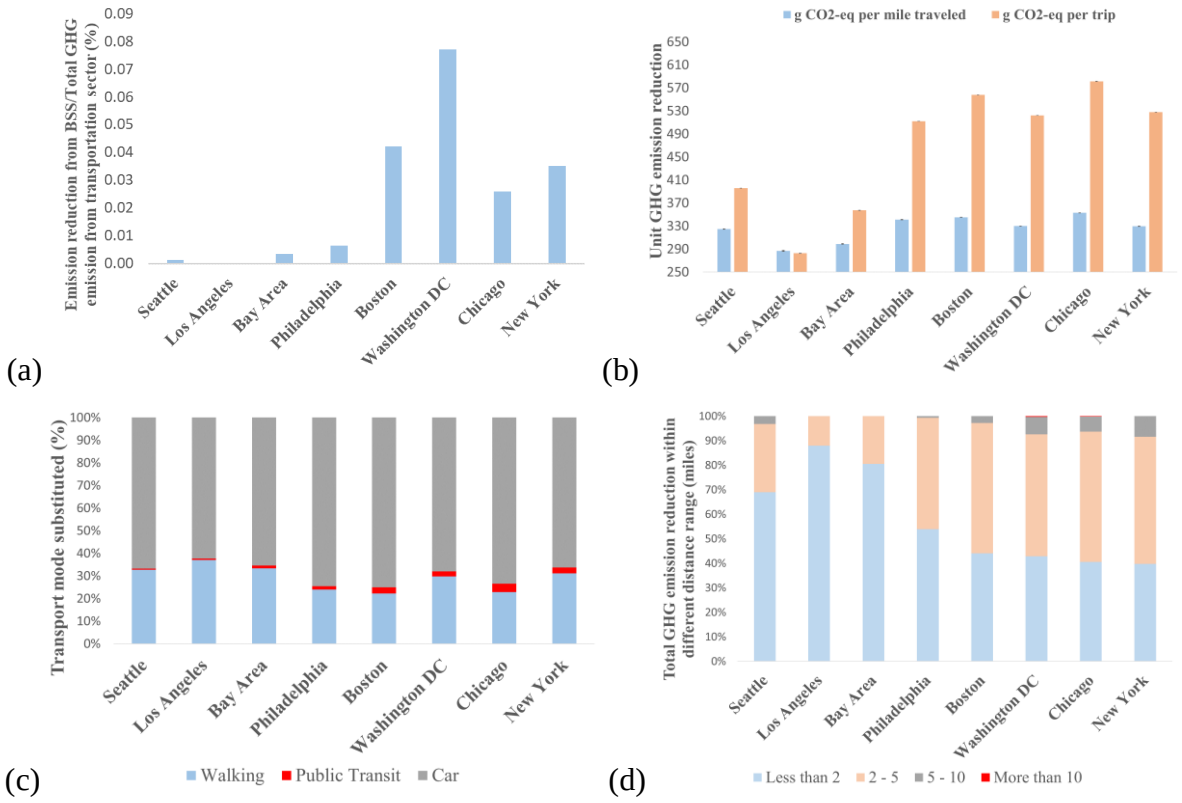


Figure 3.5. GHG emission reduction in the eight cities: (a) Emission reduction from BSSs/Total GHG emission from transportation sector (%) (b) emission reduction per mile travelled and per trip, (c) percentage of trips replaced by bike share, (d) emission reduction of BSSs in different distance range

3.4.2 Spatial distributions of GHG emission reduction

To better understand how locations of the bike share stations affect the GHG emission reductions, this study also analyzed the spatial patterns of station level emission reduction (to save the computation time, the analysis in this part is based on the trip data in August, 2016 instead of the entire year's data). This information could potentially help the decision making in siting future bike share stations. The GHG emission reduction also exhibits geographic variance within a bike share system. This study allocates the GHG emission reduction of each trip to its start station (since the origin of a trip generally reflects the demand of travel) and plotted the GHG emission reduction by station on the map. In the case of Los Angeles and Philadelphia, stations located in the center of the city reduced more GHG emissions (Figure 3.6 (a1) and (b1)), because of the larger ridership (count of trips) at these stations. The GHG emission reduction per trip exhibits the reverse pattern: each trip reduced less GHG emission if they started from the city center (Figure 3.6 (a2) and (b2)). This is because trips originated from the city center are mostly short and more likely to replace walking, while trips made from areas away from city center are relatively longer. Figure 3.6 (a3) and (b3) show that higher proportion of bike share trips replaced vehicle trips in the stations that are away from the city center. Generally, all the eight cities show similar spatial patterns regarding GHG emission reduction (see Figure B. 1 in the APPENDIX B for the figures of the other cities). These results show that, although the city center has more BSS users than areas away from city center, a higher proportion of these users adopt bike share as a substitution of walking. In order to improve the environmental benefits from the BSSs, BSS operators needs to either attract more vehicle users in the city center to switch to bike share, or strategically locate and install more stations in the areas that are away from the center of the city.

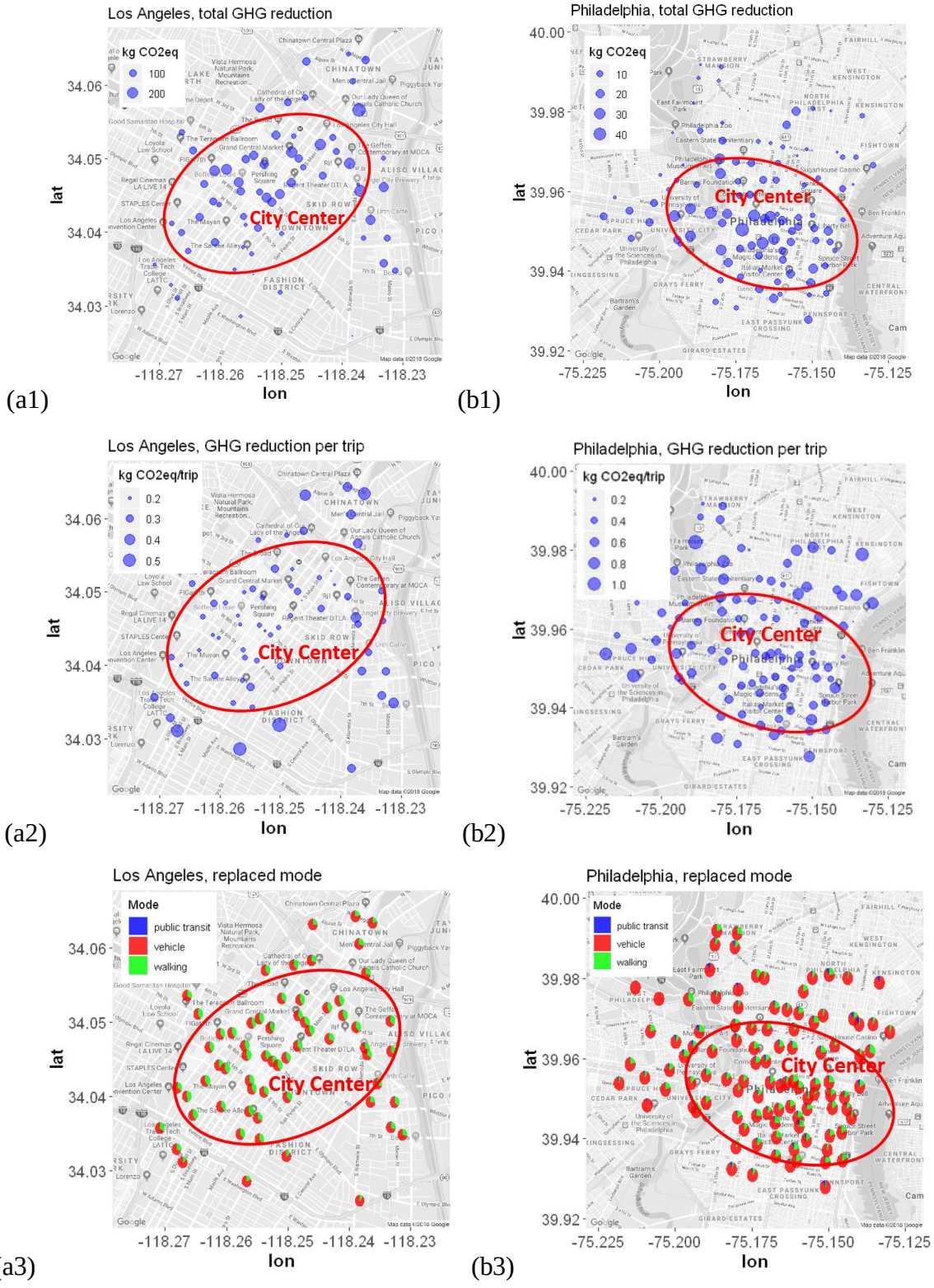
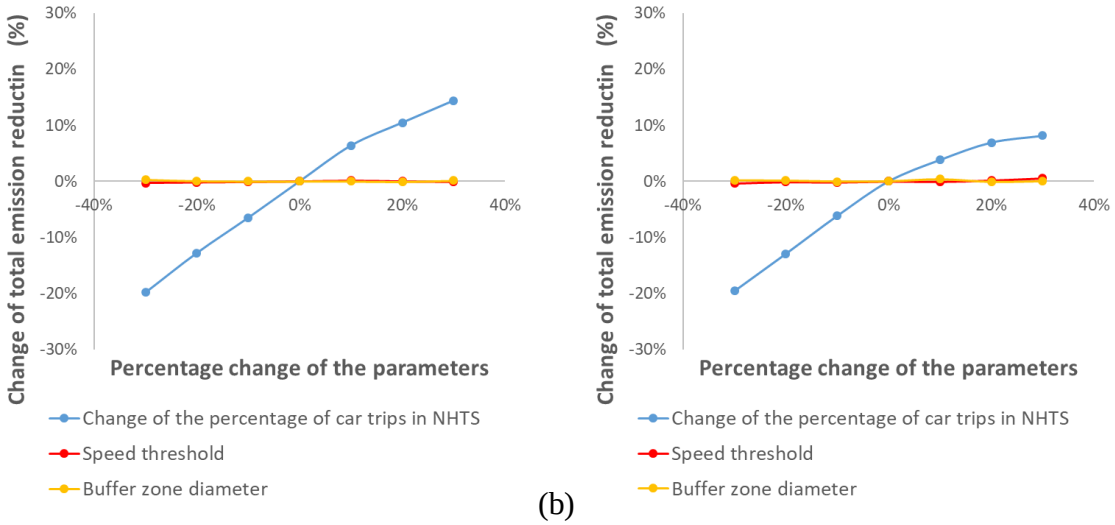


Figure 3.6. Geographic distribution of GHG reduction and mode substitution in (a) Los Angeles and (b) Philadelphia: 1. Total emission reduction of each station in Aug. 2016, 2. Emission reduction per trip, 3. Pie chart of mode substitution of each station

3.5 Sensitivity Analysis

In this model, there are several major factors or assumptions that may impact the results: speed threshold to determine the trip purpose, buffer zone radius for the public transit accessibility, transportation mode distribution from the NHTS data, and the emission factors used in the analysis. Therefore, this study conducted a sensitivity analysis to evaluate how sensitive the results are to these factors. For the speed threshold, buffer zone radius, and transportation mode distribution, the parameters are varied from 70% to 130% of the values used in the original model, and compared the total emission reduction from the BSSs in Bay Area and Philadelphia (Figure 3.7). For the emission factors, the lowest and the highest values are used for vehicle and public transit in Table 3.1 to generate the lower and upper bounds for the emission reduction. Additionally, since the emission factor of bicycling in (Dave, 2010) does not consider the emissions generated from bike share station (including docks) manufacturing and bike rebalance, this study also add the unit emissions from bike share stations and rebalance from the work of Luo et al. (Luo et al., 2019) to evaluate the influence from station infrastructure and operation.



(a) (b)

Figure 3.7. Change of total emission reduction by varying parameters in the model: (a) Bay Area, (b) Philadelphia

In the original model, the average trip speed was used as the speed threshold to determine whether a trip is for commuting or leisure purposes. So the sensitivity analysis tested the result variations using a speed threshold that ranges from 70% to 130% of the average speed in each city (as listed

in Table 3.3, the average speed is different in each city). A buffer zone radius of 200 meters was originally used to extract the public transit accessibility information. Accordingly, the sensitivity analysis tested the result variations using a buffer zone radius ranging from 140 meters to 260 meters. For the distributions of transportation mode choices, the probability of the car trips is modified by -30% to +30% of the original value, and then evenly distributed the changes to walking and public transit (for example, if the mode distribution for a trip at given distance is 60% by car, 30% by walking, and 10% by public transit, the -30% scenario will decrease the car trip probability to 42% and increase the probability of walking and public transit to 39% and 19%, respectively). These scenarios will represent cases where the bike share users take less or more vehicle trips than an average person. The focus of these different scenarios is on vehicle trips because the vehicle trips dominate the potential GHG emission reduction (Figure 3.5(c)). For the emission factors, this study used the travelled mileage as the weights to calculate the weighted average for vehicle and public transit. In this part, instead of using the weighted average, the lowest value (sedan – 382 g $CO_{2\text{-}eq}$/passenger mile travelled (PMT) was used for vehicle, and BART – 136 g $CO_{2\text{-}eq}$/PMT for public transit) and the highest value (pickup – 619 g $CO_{2\text{-}eq}$/PMT was used for vehicle, and Bus average – 326 g $CO_{2\text{-}eq}$/PMT for public transit) to calculate the emission reduction in two extreme cases (Table 3.4) In addition, in order to evaluate the influence of the emissions from stations and rebalance, this study adds the unit emission of 95 g $CO_{2\text{-}eq}$/PMT (from Luo et al. (Luo et al., 2019)) for bike share stations (57 $CO_{2\text{-}eq}$/PMT, including docks) and rebalance (38 $CO_{2\text{-}eq}$/PMT) to the original emission factor of 33 g $CO_{2\text{-}eq}$/PMT (from Dave (Dave, 2010)) for the emissions from bike, road and human breath, which results in a modified emission factor of 128 g $CO_{2\text{-}eq}$/PMT for bicycling using shared bikes from station-based BSSs. It is notable that the emission factors from Luo et al. (Luo et al., 2019) reflect the average operational conditions of a station-based system in the U.S.. Because the modified bicycling emission factor is the sum of emission factors from two studies, which may rely on different data source and assumptions, this study only considers emissions from station and rebalance in the sensitivity analysis. In the sensitivity analysis, this study uses the same emission factors of walking, public transit, and vehicles as stated in Table 3.1.

Table 3.4. Change of total emission reduction by varying the emission factors for public transit, vehicles and bycycling

Change (%) of total GHG emission reduction	Seattle	Los Angeles	Bay Area	Philadelphia	Boston	Washington DC	Chicago	New York
Using lower bound of the emission Factors for public transit and vehicles	-14.7%	-14.7%	-15.0%	-15.2%	-15.7%	-15.6%	-15.9%	-15.7%
Using upper bound of the emission Factors for public transit and vehicles	43.2%	43.3%	42.8%	42.5%	41.8%	41.9%	41.4%	41.7%
Using modified emission factor for bicycling	-29.3%	-33.1%	-31.8%	-27.9%	-27.5%	-28.8%	-26.9%	-28.8%

The sensitivity analysis shows that the total emission reduction is the most sensitive to the emission factors and mode choice distribution change, but stays relatively stable when changing the speed threshold and buffer zone diameter. Therefore, it is critical to obtain more accurate information about BSS users' transportation choices to refine the estimated environmental benefits of BSSs. It is notable that, for longer trips (e.g., over 5 miles), driving is the dominant mode (e.g., with a probability of 90%). When evaluating the +30% scenario, the probability of driving would exceed 100%. In these cases, this study capped the probability of vehicle trips to be 100% and all trips would replace vehicle trips). This is why the line for the mode distribution change is not linear in Figure 3.7 (the positive side is curved due to the capped 100% probability). When only considering changing the percentage of vehicle trips by -30% to 0% and fit a linear regression, the linear trend line has a gradient of 0.6567 in Bay Area (R squared 0.9994) and a gradient of 0.6542 in Philadelphia (R squared 0.9996). These results show that, if the bike share users take 10% less vehicles trips than an average person (whose travel pattern is captured in the NHTS data), the total emission reduction of the system will decrease by 6.6% and 6.5% in Bay Area and Philadelphia, respectively.

When the emission factors are changed in the analysis (Table 3.4), the total emission reduction of all the studied cities decreased by 14-16% for the low emission factor case and increased by 41-43% for the high emission factor case. This provides a lower and upper bound of the total emission reduction given the uncertainty of the emission factors in different cities. It is worth mentioning

that it is unlikely to expect the high emission factor case to happen, which assumes that all the replaced vehicle trips used pickups. When using the modified emission factor for bicycling (adding emissions of stations and rebalance from Luo et al. (Luo et al., 2019)), the total emission reductions are lowered by around 30% (Table 3.4). Therefore, the emissions from bike share stations and rebalance operation also play an important role for the environmental performance of a BSS.

3.6 Conclusions and Limitations

This study proposed a Bike Share Emission Reduction Estimation Model (BS-EREM) to evaluate the environmental benefits of bike share systems in eight cities in the United States. The BS-EREM model considers the trip distance, trip purpose, trip start time, public transit accessibility around each bike share station, as well as the historical distributions of transportation mode choices in different cities. Although the analysis only focused on the reduction of GHG emissions, the same approach can be applied to quantify other environmental impacts (e.g., NOx, PM2.5), using appropriate emission factors as inputs.

The total annual GHG emission reduction in the eight systems evaluated in this study shown linear relationships with the overall system size and ridership. According to the modeling result, as the largest bike share program in the United States, New York's Citi Bike Share contributed to 5,417 tons of GHG emission reduction in year 2016. In contrast, Seattle's BSS only reduced 41 tons of GHG emissions in 2016. However, the emission reduction contribution from BSSs currently is still relatively limited (less than 0.1%) in the context of the total GHG emissions from the entire transportation sector in these cities. Expanding the BSS system size (e.g., building more stations and docks, launching more bikes) can help increase the GHG emission reduction by generating more bike share trips. Another thing to note is that, for Los Angeles, even though the overall traffic volume is at the top level in the U.S. (Gold et al., 2015), the usage intensity (trips per bike, per station, and per dock) of its bike share system is relatively low, which means that bike share is not as popular there as in New York City. Improving biking infrastructures such as bike lanes can potentially help improve the use of bike share (United Nations Environment Programme (UNEP), 2018).

In addition, the environmental benefit is also influenced by the mode substitution, i.e., which transportation mode does the bike share trip replace. In the analysis, the bike share is mainly replacing trips made by vehicles and walking, while the distribution of replaced modes varies among different cities. The percentage of bike share trips replacing car trips is higher in Philadelphia and Chicago, compared to other cities, leading to higher GHG emission reduction per mile travelled. Such variation is mainly due to the difference of transportation mode choice preferences in these cities and the spatial coverage of the BSSs, which impacts trip distances.

Furthermore, the bike share system layout matters. The GHG emission reduction is mainly contributed by short bike share trips (less than 5 miles). Bike share stations in the center of a city generally have higher total GHG emission reduction but lower unit GHG emission reduction (per trip travelled). The replaced transportation modes also exhibit spatial variations: a higher proportion of bike share trips replaced car trips in areas further away from city center than in the downtown area. Therefore, the unit emission reduction potential (per mile) from BSSs could also be influenced by its bike share station network (i.e., how many stations are located in downtown versus areas away from city center, in the business centers versus leisure areas etc.). In order to improve the unit environmental benefit of bike share (per trip or per mile travelled), the system operators need to attract more vehicle users in the city center to switch to bike share, and also attract more users in the areas away from the city center. In order to achieve these goals, the cities need to improve biking safety (for instance, having more bike lanes) and the convenience of using bike share (such as simplifying the process of checking out and returning bikes), which would make bike share a more competitive transportation mode. The stations that are further away from city center are often distributed less densely than in city center, which can be observed in Figure 3.6. The difficulty of finding a nearby bike share stations in areas away from city center may discourage users to choose bike share. Installing more stations in these regions could help increase bike share usages but will increase the emissions from the station manufacturing. Having a dockless system in areas with less demand can reduce the need of stations and the associated emissions but may increase emissions from rebalancing dispersed bikes. Future study considering these factors at the station level is needed to further quantify the tradeoffs and assess the net benefits of different approaches. In addition, as indicated by Luo et al. (Luo et al., 2019) and the sensitivity analysis results, improving the rebalance efficiency can further improve the

sustainability of a BSS. This can be achieved by optimizing the station location and capacity (García-Palomares et al., 2012; Park & Sohn, 2017; Romero et al., 2012), applying more efficient rebalance vehicle routing algorithms (Alvarez-Valdes et al., 2016; Chiariotti et al., 2018), and using more environmental-friendly vehicles (e.g., electric vehicles) instead of internal combustion engine vehicles for bike rebalancing.

In the case of a dockless BSSs or hybrid BSSs (combining dockless and station-based BSSs in the same system), the proposed BS-EREM can still be applied. To reduce the computational intensity due to the dispersed bikes, the service area can be partitioned into small grids. Then the center of each grid can be modelled as a pseudo station. All trips starting (ending) within a grid can be marked as starting (ending) at the corresponding pseudo station. It can be checked whether there are public transit facilities within the grid to determine the public transit accessibility. The grid size should be identical to the spatial coverage of an existing bike share station. Such map gridding method has also been used in existing studies (such as in (Pan et al., 2019; Zhang & Mi, 2018)) to model a dockless system.

Although this study has the merit of proposing a more detailed model to quantify the environmental benefits of BSS programs, there are three major limitations that need to be pointed out. First, the results are based on the assumption that the BSS users have the same travel pattern as an average person in the urban area of each city. The sensitivity analysis has showed that the results are very sensitive to the transportation mode distribution. Therefore, future research collecting more detailed transportation mode distribution specifically from the BSS users can help improve the results accuracy. Additionally, other factors, such as weather and user demographics may also impact the mode substitution decisions. For example, this study only evaluated whether public transit stations and stops are available near the bike share stations in the model but did not consider the actual schedule and routes of the public transit networks. It is possible that the users may have to wait for a long time or have to make too many transfers between two bike share stations having access to public transit, making the substitution of public transit less feasible. When surveying BSS users on mode substitutions by bike share trips, it is important to focus the question on a specific trip, so mode substitution can be linked to trip distance, purpose, start time, origin, and destination, etc. Such information would also benefit the system operators and city planners to

optimize the system and attract more vehicle users to switch to bike share. Second, the emission factors used in this study may overestimate the environmental benefits of bike share trips since emissions from other system infrastructure and operations such as bike share station and bike rebalancing are not included. The sensitivity analysis showed that the emissions from bike share stations and rebalance could decrease the overall GHG emission reduction by around 30%. Emission factors from a more comprehensive life cycle assessment for different transportation modes, including bike share, tailored to each city would help improve the estimation using BS-EREM. Also, the system operators of BSSs may apply different rebalancing schedules and use different vehicles (e.g., trucks, vans, or electric tricycles). Therefore, to obtain more accurate estimation, more detailed data on system operations in different cities are required. In addition, there could be other factors affecting the emission factors, such as the increased emission from extensive usage of air conditioning in vehicles in certain seasons, different driving behaviors, and vehicle ages etc., which are not reflected in this study. Furthermore, different BSSs may use bikes and stations that are manufactured with different materials or processes and apply different maintenance and disposal practices. These factors will also affect the system's GHG emissions and require more detailed data to be further addressed. Third, this study only considers the cases that bike share can substitute the entire trip. Because the data only record trip start and end stations, there is no information on the actual trip origin and destinations. Bike share could also serve as the "last mile" solution and contribute to substituting a one-mode trip (e.g., car trip) with a multi-modal trip (e.g., bike-train-bike mobility chain as mentioned in (United Nations Environment Programme (UNEP), 2018)). Due to the lack of data to support such analysis, the multi-modal trip substitution is not considered, which could underestimate the environmental benefits of bike share. Data that tracks trips across different transportation modes (e.g., smart cards that can be used for both public transit and bike share) and more targeted survey questions could help further analyze these trips.

4. ANALYZING STATION INTERACTIONS AND IMPROVING DEMAND PREDICTION FOR BIKE SHARE SYSTEM EXPANSION

4.1 Introduction

The analysis in Chapter 3 shows that BSSs can reduce carbon emissions of urban transportation systems. To better take advantage of such environmental benefits and other benefits of BSSs, system operators are implementing different strategies to improve the BSSs. One of the most important strategies is to expand the system. As shown in Figure 1.2, many cities kept adding new stations to their station-based systems to expand the overall spatial coverage of the system or increase the density of stations at high demand areas. Figure 4.1 shows the locations of the BSS stations in Chicago as well as their year of launch (i.e., the year when the stations first appeared in the trip records), illustrating the progress of expanding the station network to the outskirt areas and also increasing the station density in the central areas.

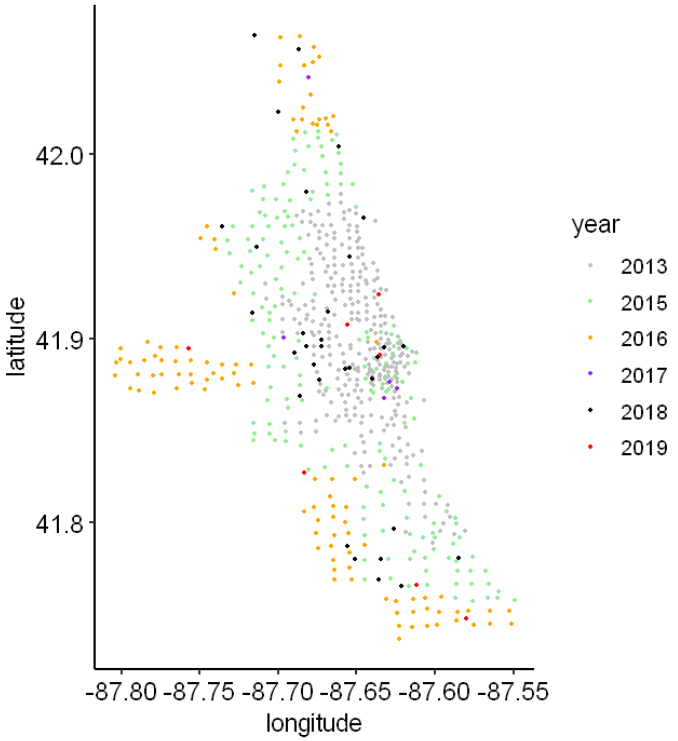


Figure 4.1. Locations of the stations in Chicago and their launch years

When planning the expansion of the station network, it is crucial to estimate the demand of the newly added stations, which can guide the decision of choosing the station capacity and allocating bikes (Zhang et al., 2016a). However, very few studies have modeled the demand prediction for BSS expansion. The major challenge for such prediction problems is that there are no historical trip records for the new stations, thus lacking information to predict the demand for new stations. Therefore, existing studies highly rely on external information such as socio-demographic and Point-of-Interest (POI) attributes to make the predictions (Liu et al., 2017; Watson & Telenko, 2019). Such external data requires time and effort to collect and process; in addition, because different models used different data and the data availability may differ across cities, the model transferability is often limited when applying the models to different cities. Besides, these studies also only focused on those new stations but did not consider that new stations may also influence the demands of existing stations (more details about related literature are discussed in Section 4.2).

However, some empirical studies have shown that there exist interactions between the new stations and existing stations (Hyland et al., 2018; Zhang et al., 2016c). New stations may compete with existing stations regarding the demand when they are very close; while new stations could also complement existing stations and increase their demands since the new stations increase the choices of locations to pick up or return bikes, thus encouraging more usages at some existing stations. Considering such station interactions, the demand prediction for BSS expansion should not only focus on the new stations, but also account for the resulted demand change of existing stations. Models ignoring the station interactions may lead to incorrect decisions when allocating docks and bikes. Therefore, demand prediction models that have low dependence on external data and also consider the station interactions in the BSS expansion process are greatly needed.

In addition, although existing studies observed the station interactions in the system expansion process, the understanding of the station interactions is limited. For example, Hyland et al. (2018) have pointed out that stations compete with each other when their distance is within 0.5 miles, but complement each other when the distance is between 0.5 - 3.1 miles. However, the conclusion is drawn based on an arbitrarily selected threshold (0.5 miles); the study also did not analyze whether different distances within the same range (e.g., 1 mile and 2 miles in the range of 0.5 - 3.1 miles) impact the station interactions to different extent (there are more discussions about relevant

literature in Section 4.2). Therefore, a more detailed analysis is needed to reveal how the station interaction effects change with the distance between stations. Such knowledge not only helps the modeling of BSS expansion demand predictions, but also brings insights to the design of station networks in BSSs. For instance, the station networks should take advantage of the complement effects between stations, and also avoid the competitions between stations that are too close.

This study aims to improve the understanding of interactions between new and existing stations and better model the bike share demand in the BSS expansion process. To achieve these objectives, a piecewise regression is first conducted to analyze the station interactions, which provided the detailed mathematical formulation of how the demands of existing stations change corresponding to their distance to the nearest new stations. Then based on the observations from the piecewise regression results, features of spatial station density in fine-grained concentric bands are integrated into BSS expansion demand prediction models, which improves the prediction by considering the stations' interactions with other stations in different distance ranges. A Spatial Eccentricity Quantile based Ensemble Model (SEQEM) is proposed that requires no external data but yields better prediction performance than the model using external data.

The rest of this chapter is organized as follows: Section 4.2 summarizes the related literature, identifies the research gap, and discusses the contributions of this study. Section 4.3 introduces the data and methodology. The results are presented in Section 4.4. Lastly, Section 4.5 concludes the findings and discusses the limitations of this study and potential future work.

4.2 Literature review

Demand estimation is a crucial component of BSS planning and system operations. For system expansion, planners normally estimate the steady-state demands of the stations (e.g., average daily demands in a month) (Watson & Telenko, 2019). For example, using the demand levels and capacities of existing stations as a reference, the planners can determine the capacities of new stations based on the estimated demand level, assuming that the capacities should be proportional to the demands (Institute for Transportation & Development Policy, 2018). Regarding operational issues such as bike fleet rebalance, more dynamic predictive models are required. In order to

prepare for the truck rebalancing scheduling, the operators need to know the demand for a specific station/area in a short time interval such as one hour (Caggiani et al., 2018; Chiariotti et al., 2018).

So far, the majority of existing studies focus on the short-term demand prediction. In such studies, the historical demand data for each station or area is available to help the models learn the demand patterns and predict the upcoming short-term demands (Chen et al., 2016b; Jia et al., 2019; Li et al., 2015, 2016; Yang et al., 2016; Zhou et al., 2019). Meteorology and event data is often integrated into the models to reflect the impact of extreme weather and special events such as concerts (Chen et al., 2016a; Xu et al., 2018a, 2018b).

Compared with the short-term BSS demand prediction models, there are a limited number of studies focusing on the demand estimation for system expansion. One key difference between shot-term prediction and expansion prediction is that there is no trip data for the new stations. Therefore, existing studies have to depend on external data such as local socio-demographic and point of interest (POI) information around the stations to make the predictions for new stations. Watson and Telenko (2019) built a binomial parameter mapping model to predict the demand for station expansion by integrating the population size (at the level of 0.25-mile spatial grids) and user population product affinity (defined as the ratio of bike share usage to population). Liu el al. (2017) first clustered the stations into functional zones by considering the POI characteristics and close geographical distances; then estimated the demands at the zonal level using information about trip distance preference, zone-to-zone preference, and zone characteristics. Then the zone-level demands are distributed to the stations within the zone based on the POI information around the stations. There are two limitations in these studies. First, these studies highly rely on the external socio-demographic and POI data. Such external data could be time consuming to collect and process; the used data variables also differ across different studies, which makes it difficult to generalize the models to other cities. Second, these studies only reported the prediction performance on new stations and ignored the station interactions in the BSS expansion process. The station interactions (Hyland et al., 2018; Zhang et al., 2016c), i.e., the effects that new stations could compete with or complement the existing stations regarding the demand, impact the demand of existing stations and thus the demand of the whole system. For instance, Liu et al. (2017) separated the BSS service area into different functional zones and assumed that adding new

stations has little impact on the zone level demand. However, in the case that the new stations complement the demand of existing stations due to improved spatial service coverage, their method would underestimate the demand of the whole zone.

Although some studies have pointed out the existence of the abovementioned station interactions in their empirical analysis, there still lacks detailed understanding about at which distances the stations compete with or complement each other. Wang and Lindsey (2019) analyzed the impact of changing station networks on the bike share usage of annual members (registered long-term users). They adopted a quasi-experimental, difference-in-difference approach to analyze the members' bike share trip data from 2010 to 2015 in Minneapolis-St. Paul, Minnesota. Their analysis revealed that members were more likely to increase their usage of bike share when the distances from their homes to a station decrease. Similarly, Zhang et al. (2016) analyzed the changes of usages from new users and steady users in the expansion process in the BSS in Zhongshan, China. They found that adding new stations in high-demand areas can attract new users and encourage the usage of steady users, while new stations could also compete with old stations. However, the detailed relationship between the station density and the demand change pattern is not analyzed. Hyland et al. (2018) pointed out a distance threshold for the station competition and complement relationship. They proposed a cluster stations and regress (CSR) approach to model the usage of stations in a BSS, which combined station clustering and multilevel mixed-effect regression models. In their regression model, the number of nearby stations within 0.5 miles (0.8 km) and the number of other stations within 0.5-3.1 miles (0.8-5 km) to a certain station are used as the variables to model the station usage. By observing the sign of the coefficients of these variables, they found that station usage increases when the number of stations within 0.5-3.1 miles increases, while the station usage decreases when there are more stations within 0.5 miles. However, it is unclear by how much the distance will impact the demand for stations in the same distance range. For example, adding a new station in 1 mile or 2 miles away from an existing station may impact the existing station's demand by different magnitude, although 1 mile and 2 miles are both in the range of 0.5-3.1 miles. Also, their choice of the 0.5 miles (0.8 km) distance threshold is arbitrary, which may not capture the real threshold where the stations' relationship changes from competing to complementing.

In summary, existing studies have not analyzed the detailed relationships on how the station interactions change with the distance between the stations. Previous demand prediction models for the BSS expansion also did not consider the station interactions' impact on existing stations. Existing BSS expansion demand prediction models also highly rely on the external socio-demographic and POI data, which requires a significant effort to collect and process. This study contributes to the existing literature in the following aspects: (1) the mathematical relationship between the demand changes of existing stations and their distance to the nearest new stations is revealed using a piecewise regression model, which improves the understanding of the station interactions in the expansion process; (2) features of spatial station density in fine-grained concentric bands are constructed to represent the other stations in different distance ranges that a certain station can interact with, which are then integrated into the expansion demand prediction models; the case study on the BSS in Chicago shows that integrating such network spatial structure information improves the prediction; (3) A Spatial Eccentricity Quantile based Ensemble Model (SEQEM) is proposed, which does not require external socio-demographic and POI data but yields better prediction performance than models using the external information.

4.3 Data and method

Notation list:

a_{iym}	The number of days that the station i is active in the month m of the year y;
c_{iym}	Total bike withdrawals at station i in the month m of year y;
d_{iym}	The distance from station i to the nearest new station in month m of year y;
EV_Q	The evaluated model performances corresponding to the specific Q
D_{iym}	The band distance from station i to the nearest new station in month m of year y; The band distance (in miles) takes the value of the mid-point of the corresponding band interval, so $D \in \{0.05, 0.15, ...\}$;
$f_{t,}$	The forecasted value for the t^{th} instance out of all the T instances to be predicted; in the case of this study, an instance is the demand of one station (in a certain month and year) to be predicted;
$G(D)$	The aggregated demand growth rate corresponding to the band distance D;
MD_{ym}^{X}	The prediction model based on the "X" category of features, trained using data in the month m of year y; for instance, MD_{ym}^{GEO} denotes the model based on the "GEO" features, trained using data in the month m of year y;

o_t	The observed value for the t^{th} instance out of all the T instances to be predicted;
Q	Spatial eccentricity quantile threshold;
QL	A list of candidate spatial eccentricity quantile threshold Q;
w_{iym}	The observed average daily bike withdrawals at station i in the month m of year y.

4.3.1 Data processing

This study aims to analyze how the new stations impact the demand of existing stations. Therefore, the demand information needs to be first extracted from the bike share trip data. This work selects Chicago as the case study city because the station-based BSS in Chicago is a desirable example that keeps expanding the system over the years. As shown in Figure 4.1, the BSS in Chicago was first launched with 300 stations in 2013, and with new stations added to the system every year (except 2014), the BSS had expanded to 610 stations by the end of 2019.

For the planning of BSS expansions, the decision makers care more about the steady-state demand level rather than the short-term (e.g., hourly) demand fluctuations (Hyland et al., 2018; Watson & Telenko, 2019). This study focuses on the bike withdrawals from the stations, which are consistent with most of the previous studies modeling bike share demand (Hyland et al., 2018; Li et al., 2016; Lin et al., 2018; Singhvi et al., 2015; Wang & Kim, 2018), but the same models can also be applied to the bike returns. The demand variable of interest is the average daily bike withdrawals w_{iym} at station i in the month m of year y. The average daily demand is computed instead of the total monthly demand to account for the fact that a station may not be available for all the days in a certain month. For example, one station was added to the system in September 16, 2016 and was only active for 15 days in that month. In this case, when evaluating its demand change between 2016 and 2017, it is unfair to compare the total bike withdrawals in 30 days in September 2017 with the total bike withdrawals in 15 days in September 2016. Therefore, the w_{iym} is computed as:

$$w_{iym} = \frac{c_{iym}}{a_{iym}} \tag{4.1}$$

where c_{iym} is the total bike withdrawals at station i in the month m of year y, and the a_{iym} is the number of days that station i is active in the month m of the year y. The first available date of each station is identified as the date that the station first appeared in the trip record. Regarding the

demand change of a certain station, this study always compares a station's demand in the same month between two consecutive years. A station is marked as "new" for a specific month if it was available in this month but was not active yet in the same month of the previous year.

Besides staying consistent with the previous study (Hyland et al., 2018), this work aggregates the data by month instead of year or day also for the following reasons: (1) there are significant demand variations between different months in a year, thus comparing the average daily demand of the whole year for a certain station could bias the result if a station is not active for the entire year. For instance, Figure 4.2 shows the average daily bike withdrawal of each month for one station (id 359). The demand was high in the middle of the year (June – September) and was at a much lower level in the meteorological winter (December, January, and February). Such demand variation pattern is common for the entire system due to the cold winter weather in Chicago (Zhou, 2015). The station 359 was added to the system in May 2015. When computing the average daily bike withdrawals of this station in 2015, it may overestimate the value for 2015 when comparing it with other years since the station is not available in the four low-demand months (January – April) in 2015. (2) Rather than analyzing the total demand in a specific day, the data is aggregated in each month because the demand in a specific day could be impacted by many external factors, such as special events or extreme weather. Aggregating the whole month's data can reduce such uncertainty.

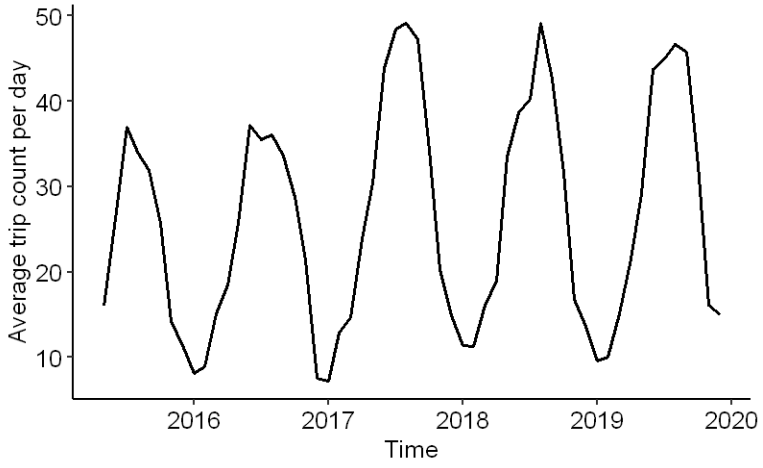


Figure 4.2. Average trip count (bike withdrawals) per day in each month for one station (id: 359)

4.3.2 Piecewise regression

In order to analyze the relationship between the demand changes of existing stations and their distance to the nearest new station, the next step is to compute the distance. Let d_{iym} denote the distance from station i to the nearest new station in month m of year y. Note that all distances in this study are computed using the great-circle distance to stay consistent with the most relevant study from Hyland et al. (2018) which also analyzed the data of Chicago's BSS, so the results of this work can be compared with their study. Then the d_{iym} of different stations is separated into a series of distance bands at 0.1-mile interval: 0 – 0.1, 0.1 – 0.2, …, which are used to group stations that have similar d_{iym} values. Depending on which band the d_{iym} value lies in, the station i in the month m of year y is assigned a band distance D_{iym}. The band distance D takes the value of the mid-point of the corresponding band interval, so $D \in \{0.05, 0.15, ...\}$ (some examples of computing the band distances are illustrated in Figure 4.3). Then the aggregated demand growth rate $G(D)$ in each band is computed using Eq. (4.2), which is the total demand of all the stations i that have $D_{iym} = D$, divided by the summation of their corresponding demand in the previous year.

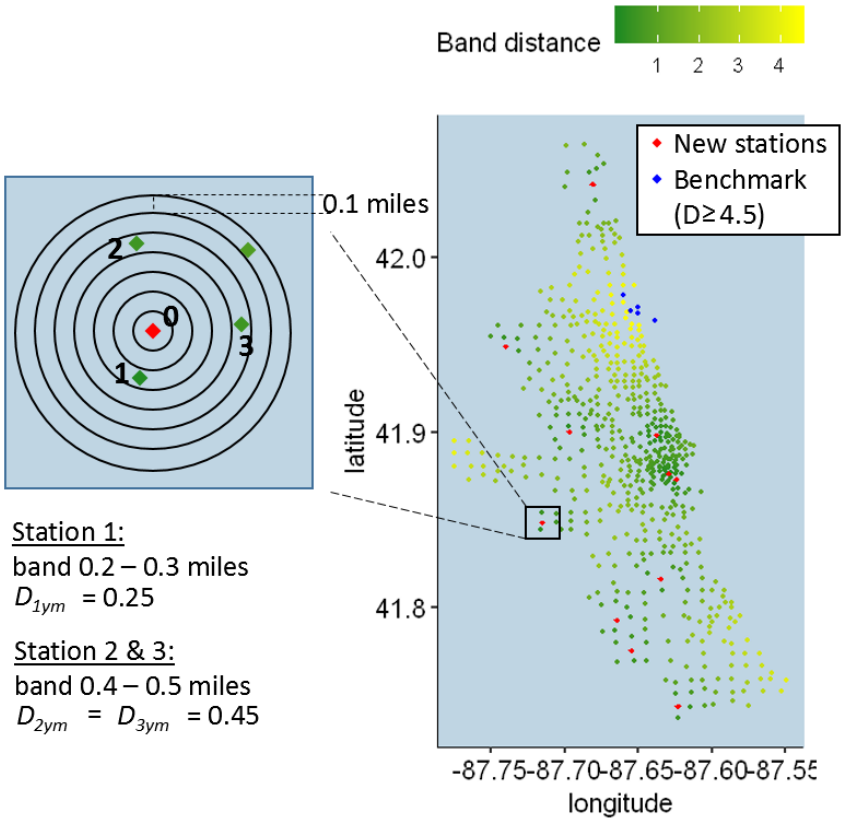


Figure 4.3. Spatial distribution of the band distance (in miles) in August 2017 and an example to compute the band distance (new stations are colored in red, and benchmark stations, i.e., existing stations that are more than 4.5 miles away from the nearest new stations, are colored in blue): station 0 is the nearest new station for station 1, 2, and 3, since station 1 lies in the band of 0.2-0.3 miles, it is assigned a band distance of 0.25 miles (the middle point of 0.2 and 0.3); similarly, since station 2 and 3 are both located in the band 0.4-0.5 miles, they are assigned a band distance of 0.45 miles.

$$G(D) = \frac{\sum_i \sum_y \sum_m [w_{iym} \mid D_{iym} = D]}{\sum_i \sum_y \sum_m [w_{i(y-1)m} \mid D_{iym} = D]} \tag{4.2}$$

Such aggregation is applied to reduce the impact of demand noises from some stations with low demands. For example, station 193 had average daily bike withdrawals of 0.17 and 0.7 in the month of January in 2015 and 2016, respectively. If only computing the growth rate for itself, this corresponds to a large growth rate of 4.1. However, the absolute demand change was only 0.53

bike withdrawals per day, which had very limited contribution to the demand changes of the whole system. In the next year, 2017, its demand rate retuned to 0.17 bike withdrawals per day. Therefore, for such low-demand stations, multiple extra trips caused by some special events in a certain month can easily lead to a very large demand growth rate, which could bias the analysis results (some other examples of such demand noises are exhibited in Section C.1 in the APPENDIX C). In contrast, for a high-demand station that has a daily bike withdrawals rate of 100, a demand increase of 20 bike withdrawals per day can only lead to a growth rate of 1.2. When aggregating the demand of all the stations in the same distance bands, the impact of noises from the low-demand stations can be reduced.

In this study, growth rate *G(D)* is only computed for each band when *D* is less than 4.5 miles. This threshold is selected based on the trip distance of historical trips. In all the trips from 2013 to 2019, the travel distances of 99% of the trips are within 4.5 miles. This means that, when the nearest new station is more than 4.5 miles away from a certain existing station, the probability that there is a trip between the two stations is very low (less than 1%), thus the new station has nearly no impact on the existing station. Therefore, the aggregated growth rate for such stations is computed using Eq. (4.3) as a "benchmark" growth rate $G_{benchmark}$, which represents a natural growth rate when the existing stations are not impacted by new stations. The growth rate *G(D)* for different band distance *D* will be compared with this $G_{benchmark}$ to learn about the impact from the new stations.

$$G_{benchmark} = \frac{\sum_i \sum_y \sum_m [w_{iym} \mid D_{iym} \geq 4.5]}{\sum_i \sum_y \sum_m [w_{i(y-1)m} \mid D_{iym} \geq 4.5]} \tag{4.3}$$

Next a piecewise regression model is fitted to analyze the relationship between *G* and *D*. Piecewise regression is also referred to as change-point regression, segmented regression, and broken-line regression (Kim et al., 2000; Ratcliffe, 2012). A piecewise regression model identifies the change points in multiple linear regression models that have continuous segments, where any of the two continuous segments change the directions at the corresponding change point (Hinkley, 1971). This study considers piecewise regression because it was applied in another study by Ratcliffe (2012) that also focused on the impact of distances, which analyzed the relationship between

violent crime densities and the distance from the crime spots to the nearest bars. In addition, in the Chicago data, it is observed that the aggregated growth rate *G* first increases and then decreases when the band distance *D* increases, and the trends of the two segments are near linear (more detailed results are in Section 4.4.1), which can be well modeled by piecewise regression (the analysis in Section C.2 of the APPENDIX C shows that the piecewise regression model provides a better fit to the Chicago data than a simple linear regression model). Since there is only one change point in this case, the piecewise regression model has the form of Eq. (4.4):

$$E[G \mid D] = \beta_0 + \beta_1 D + \delta_1 (D - \tau_1)^+ \tag{4.4}$$

where τ_1 is the estimated change point; β_0, β_1, and δ_1 are other parameters to be estimated. The superscript plus symbol ($^+$) indicates that $(D-\tau_1)^+ = (D-\tau_1)$ when $(D-\tau_1) > 0$, and zero otherwise.

4.3.3 Demand prediction for system expansion

Based on the observations from the piecewise regression analysis, the distances between stations contain very important information about station interactions. Inspired by this observation, this section integrates the distance information into the demand prediction model to improve the prediction performances. Specifically, features of spatial station density in fine-grained concentric bands around a station are constructed to represent the number of stations in different distance ranges that the station can interact with, which are then used to build demand prediction models.

4.3.3.1 Model task

The task of the BSS expansion demand prediction is that, given the dependent variable $\{w_{iym} \mid y = Y, m = M\}$ – the average daily bike withdrawals for each station *i* in month(*M*) of year(*Y*), train a demand prediction model using some independent variables $\{v_{iym} \mid y = Y, m = M\}$ associated with the stations. Then the model is applied to predict the average daily bike withdrawals $\{w_{iym} \mid y = Y+1, m = M\}$ based on the independent variables $\{v_{iym} \mid y = Y+1, m = M\}$ of all the stations in the same month *M* in the next year (*Y*+1). Note that there are newly added stations in year *Y*+1.

4.3.3.2 Features

The features (independent variables) considered in the demand prediction model can be classified into the following categories in Table 4.1:

Table 4.1. Categories of features used in this study

Feature category (acronym)	**Feature details**
DENS	Station density (proposed method). For each station, first construct the concentric distance bands around the station at 0.1-mile interval, i.e., 0 – 0.1, 0.1 – 0.2, …, 4.4 – 4.5 miles, then for each distance band, compute the number of other stations that are located in the band. Therefore, 45 variables will be created for each station; each variable indicates the density of stations in the corresponding band.
DENS2	Station density (literature method). Construct concentric distance bands in a similar way as "DENS", but with only two wider distance bands: 0 – 0.5 and 0.5 – 3.1 miles, which is the arbitrarily selected bands from a previous study (Hyland et al., 2018). Then compute the number of other stations that are located in the two bands.
GEO	Geographic information. It includes the latitude and longitude of the stations.
EXT	External information. It denotes the external information around the locations of the stations, which covers the following variables that are found to be important to bike share demand in the previous studies (Hyland et al., 2018; Liu et al., 2018, 2017; Médard de Chardon et al., 2017) : ***Demographic information:*** collected from American Community Survey (American Community Survey, 2017) at the census-tract level, the values of one census tract is assigned to a station if the station is located within the census tract. - Population density - Per capita income ***Point of interest:*** collected using Google Maps Places API. Such variables are the counts of points of interest (e.g., bus stations) within a 1,000 feet (305 meters) buffer around a station. The 1,000 feet is selected as an appropriate distance that people can walk between a bike share station and surrounding points of interest (Croci & Rossi, 2014; Zhang et al., 2016c). - Number of bus stations - Number of subway stations - Number of restaurants - Number of parks - Number of parking lots - Number of museums - Number of schools
ECC	Spatial eccentricity. The spatial eccentricity of a certain station *i* is defined as the average distance from station *i* to all other stations (Cazabet et al., 2017). Note that the spatial eccentricity of a certain station could be different in different months since new stations are added in the expansion process.

Table 4.1 continued

ECCQ	Spatial eccentricity quantile. For year *y* and month *m*, the ECCQ of the station *i* is the quantile value of the ECC_{iym} value of station *i* corresponding to the ECC distribution of all the stations in year *y* and month *m*.

In this study, prediction models are trained using different categories of features or a combination of different categories of features to analyze which features are more predictive for the demand prediction. The models that have been analyzed include：

- "GEO", "GEO+DENS", and "EXT"

  "GEO+DENS" model (i.e., the model trained using both "GEO" and "DENS" features) is compared with "GEO" model to evaluate whether adding the information of concentric station density helps improve the predictions.

  Because the "GEO+DENS" model performs better for stations in central areas and the "GEO" model performs better in outskirt areas of the city (discussed in Section 4.4.2.1), a Spatial-Eccentricity-Quantile-based Ensemble Model (SEQEM) is proposed to ensemble the prediction results of the "GEO+DENS" model and "GEO" model in different spatial ranges (the ranges are determined by a threshold of "ECCQ", more details are discussed in Section 4.3.3.5).

  The SEQEM is then compared with the baseline "EXT" model. The major objective of the SEQEM is to predict the demands for BSS expansion only based on the information of station locations and the spatial structure of station networks, which avoids using the external data as in "EXT".

- "GEO+DENS+ECC" and "GEO+DENS+ECCQ"

  Instead of using ECCQ as a threshold in the SEQEM, this study also directly adds the ECC and ECCQ features into the prediction models to train the "GEO+DENS+ECC" and "GEO+DENS+ECCQ" models, respectively. The performances of these two models are compared with SEQEM to evaluate the necessity of using the ensemble method of SEQEM.

- "DENS" and "DENS2"

  The prediction performances of models using only DENS and only DENS2 as features will be compared to evaluate whether the more fine-grained bands improve the predictions (Section 4.4.2.3).

4.3.3.3 Model training and hyper-parameters

This study applies the XGBoost (Chen & Guestrin, 2016) algorithm to train all the models with different features. XGBoost is used in this study because it can handle high-dimension data with feature-selection ability (Wang et al., 2019), especially for the "DENS" features that have 45 variables for all the distance bands. In addition, XGBoost can report the feature importance scores, which help identify the important distance bands for the predictions (discussed in Section 4.4.2.3). For different models, the hyper-parameters are tuned using grid search (Syarif et al., 2016) with 10-fold cross validation (more details about hyper-parameter tuning are provided in the Section C.3 in APPENDIX C).

4.3.3.4 Model performance evaluation metrics

In this study, prediction models are trained using the data in a certain month M in year Y and evaluated using the data of month M in year Y+1. Instead of only evaluating the prediction performance on new stations (Liu et al., 2017), the predictions on all the stations (including new and existing stations) will be evaluated to consider the effects of station interactions. Two metrics are adopted to evaluate the performance of predictions. The first is root-mean-square error (RMSE), which has the general form of Eq. (4.5):

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{T}(f_t - o_t)^2}{T}} \tag{4.5}$$

where f_t, and o_t are the forecasted and observed values for the t^{th} instance out of all the T instances to be predicted. In this study, an instance is one w_{iym} value to be predicted. RMSE quantifies the level of absolute errors. In this study, large absolute errors are typically observed in high-demand

stations. Another metric is the mean absolute percentage error (MAPE), which has the general form of Eq. (4.6):

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{f_t - o_t}{o_t}\right| \tag{4.6}$$

In the case of Chicago bike share data, a large MAPE error often happens at low demand stations in the outskirt areas. Therefore, evaluating the model performance using both RMSE and MAPE helps identify the poor predictions for both high-demand and low-demand stations.

4.3.3.5 Spatial-Eccentricity-Quantile-based Ensemble Model (SEQEM)

After observing the predictions of "GEO" and "GEO+DENS" model, it is found that the "GEO+DENS" model performs better in the central area of the station network (regarding RMSE), while "GEO" performs better at those outskirt low-demand stations (regarding MAPE). Therefore, to take advantage of the good performance of both "GEO+DENS" and "GEO" models in different areas, the Spatial-Eccentricity-Quantile-based Ensemble Model (SEQEM) is proposed, which switches the applied prediction models based on the spatial eccentricity quantile threshold *Q*. Besides potentially improving the overall prediction performance, another objective of the SEQEM is to identify a spatial boundary within which the station interactions should be considered. The detailed procedure of SEQEM is presented below (Algorithm 4.1):

Algorithm 4.1: Spatial-Eccentricity-Quantile-based Ensemble Model (SEQEM)

Input: $\{w_{iym}\}$ the average daily demand for all the station *i* in different months and years as well as the corresponding "GEO", "DENS", and "ECCQ" features;

QL: A list of candidate spatial eccentricity quantile threshold *Q*; in this study, $QL = \{0.02, 0.04, ..., 0.98\}$;

The model evaluation metric *Metric*, which in this study is RMSE and MAPE.

Output: $\{EV_Q\}$: The evaluated model performances corresponding to the *Q* values in *QL*

Procedure SEQEM($\{w_{iym}\}$,"GEO", "DENS" and "ECCQ" features, *QL*, *Metric*)

FOR any month *M* and year *Y* except the last year: // data in the last year is only used for testing

Train models MD_{YM}^{GEO} and $MD_{YM}^{\text{GEO+DENS}}$ using data $\{w_{iym} \mid y = Y, m = M\}$ and the corresponding "GEO" and "GEO+DENS" features, respectively

FOR *Q* in *QL*:

FOR any month *M* and year *Y* except the first year: // data in the first year is only used for training

FOR all the station *i*:

IF $ECCQ_{iYM} < Q$:

Predict the demand f_{YMi} using $MD_{(Y-1)M}^{\text{GEO+DENS}}$

ELSE: Predict the demand f_{YMi} using $MD_{(Y\text{-}1)M}^{\text{GEO}}$

Compute an overall performance EV_Q based on all the demand rates $\{w_{iym}\}$ and predicted values $\{f_{iym}\}$ using *Metric*

Note that a quantile threshold *Q* is used as the threshold instead of an absolute value of spatial eccentricity, because the spatial eccentricity of each station also changes in different months when new stations are added. When the system expands to more outskirt areas, the maximum spatial eccentricity in the station network also becomes larger. Considering this, with a quantile threshold, the relative locations can always be found (central or outskirt) for each station. Note that a smaller spatial eccentricity value means that the station is at a more central location. In this study, SEQEM is applied on all the data with a list of $Q \in \{0.02, 0.04, ..., 0.98\}$. The overall RMSE and MAPE performances are then recorded for the corresponding *Q*. By visualizing the RMSE and MAPE changes with varying *Q*, the *Q* that yields the best overall performance can be identified.

4.4 Results and discussions

4.4.1 Result of piecewise regression

First, the computed benchmark growth rate $G_{benchmark}$ equals 0.97, which is the aggregated growth rate for those existing stations that are more than 4.5 miles from their nearest new station. Therefore, the overall demand of these benchmark stations actually decreases slightly by 3%.

The fitted piecewise regression model shows a breakpoint (point A) at 0.30 miles (Figure 4.4); and the $G_{benchmark}$ (blue horizontal line) intersects with the fitted model at 0.11 miles (point B) and 4.37 miles (point C). The estimated model parameters are shown in Table 4.2. A demand dilution effect can be observed when D is less than 0.11 miles (0.18 km, 581 feet), where the new stations compete with the existing stations. This short distance is suitable for walking but it is rare that a bike share user rides between two stations that are so close. With an average human walking speed of 3.13 miles/hour (Levine & Norenzayan, 1999), this distance requires a walking time of 2.1 minutes. This distance is much shorter than the average trip distance of 1.26 miles for all the bike share trips in 2013-2019 in Chicago. It is likely that some users that previously picked up bikes at the existing stations were attracted to the new stations, which caused the demand dilution. In this case, although the new stations can provide more choices for users and increase the total station capacity in the local region, the system designer need to more carefully evaluate the trade-off between the station installation cost and the potential benefits to bike share ridership.

When D is longer than 0.11 miles, it is observed that the new stations actually complement the existing stations, leading to a higher growth rate G than the benchmark. The stations located around 0.3 miles to new stations benefited the most from such complement effects. Such complement effects can be understood as: adding new stations increases the spatial coverage of the BSS and thus the options for the users to pick up or return bikes, which may encourage more usage from existing users and attract more new users. When D is larger than 0.3 miles, such complement effects decrease when D increases.

The work of Hyland et al. (2018), which also analyzed the Chicago BSS data, concluded that adding more stations within 0.5 miles of an existing station can reduce the demand of the existing station, while more stations in the range of 0.5-3.1 miles lead to an increase of demand for existing

stations. However, since the 0.5-mile threshold is selected arbitrarily, it could happen that the observations within 0.5 miles are the mixed effects of both competition and complement between stations, where competition is more prevalent. Compared with their analysis, the piecewise regression in this study improved upon their work to provide the detailed trend of demand changes and a more fine-grained analysis of the change point.

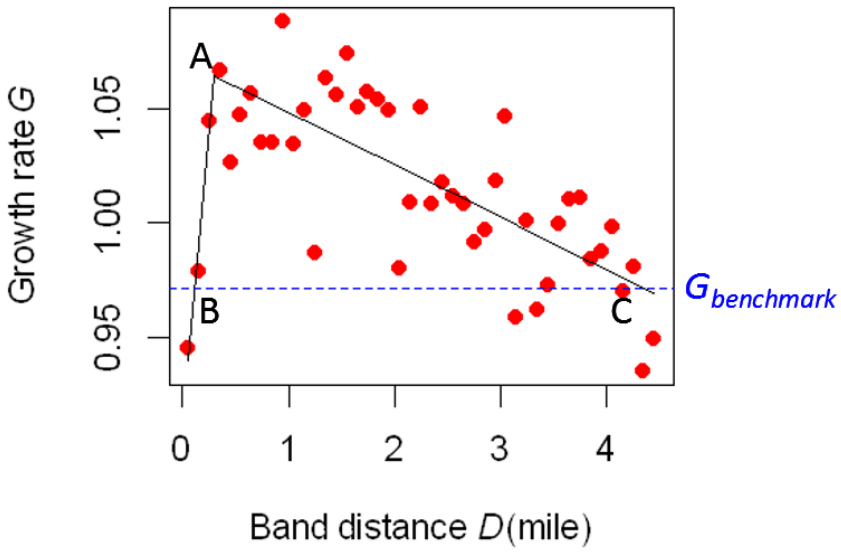


Figure 4.4. The fitted piecewise regression: the horizontal dashed line indicates the benchmark growth rate $G_{benchmark}$; point A (0.30, 1.06) is the change point; point B (0.11, 0.97) and C (4.37, 0.97) are the two intersection points of the fitted piecewise regression model with the benchmark line

Table 4.2. Estimated parameters of the piecewise regression

Parameter	**Estimation**
β_0	0.915*** (0.029)
β_1	0.495** (0.169)
δ_1	-0.023*** (0.003)
τ_1	0.301*** (0.058)
Model fit statistics	
R^2	0.624
Adjusted R^2	0.596

*Note: * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001;*

the standard errors are shown in parentheses.

4.4.2 Results of demand prediction

In this section, as outlined in Section 4.3.3.2, the overall performance of "GEO", "GEO+DENS", and "EXT" models are first compared in Section 4.4.2.1. Specifically, Section 4.4.2.1 compares the performances of the "GEO" and "GEO+DENS" models to evaluate whether integrating the spatial structure information of station networks in the "GEO+DENS" model can improve the predictions upon the "GEO" model. The spatial ranges in which the "GEO" and "GEO+DENS" models perform better than each other are also discussed. Then, Section 4.4.2.2 compares the performances of the proposed SEQEM with the baseline "EXT" model. Lastly, the performances of models using only "DENS" and only "DENS2" as features are compared to explore whether the fine-grained distance bands improve the predictions in Section 4.4.2.3. The distance ranges that are more important for demand predictions are also examined in Section 4.4.2.3.

4.4.2.1 Overall performance of "GEO", "GEO+DENS", and "EXT" models

Figure 4.5 presents the overall prediction performances of the "GEO", "GEO+DENS", and "EXT" models in different years. For all three models, the RMSE in 2015 is much higher than other years. This is because the expansion in 2015 was in a much larger scale than other years. For example, there were 174 new stations in July 2015, while in July 2014, the system only had 300 stations. In contrast, there were 82, 8, 9, and 32 new stations in July of 2016-2019, respectively. According to the locations of the new stations in July 2015 in Figure 4.6(a), the system expanded to the outskirt areas to a very large extent; in addition, a lot of new stations are also added in the downtown area, which significantly increased the station density in the downtown area. When the number of stations is quickly increased by 58%, the information learned from the last year is insufficient to provide a desirable prediction. Besides, Figure 4.6(b) shows that the stations that have large absolute prediction errors from the "GEO-DENS" model are mainly in the downtown area, where both the station density and trip demands are the largest in the system. The trained model is based on 2014 data where the station density in the downtown area is much sparser. Therefore, the "GEO+DENS" model makes predictions based on a spatial station structure that has never been observed before. Also, since "GEO+DENS" model has more features than the "GEO" model, it is more likely to over-fit to historical observations, which causes a larger error. Not surprisingly, the

"EXT" model based on external information performs better when the BSS expands the spatial coverage to a very large extent.

In year 2016 – 2019, when the expansion is more gradual, the RMSE is lower and in an identical level in different years. In these four years, overall, the "GEO+DENS" outperforms "GEO" regarding the RMSE metric (Figure 4.5(a)). However, "GEO+DENS" model has larger overall MAPE error than the "GEO" model (Figure 4.5(b)). In Figure 4.6(d), the spatial distribution of the absolute percentage errors is plotted, which shows that stations with very large percentage errors are located in the Southern outskirt area of the BSS (bottom in the figure). For example, the station 588 has a very small demand rate of 0.03 trips per day in August 2016 but the "GEO+DENS" model predicts it to be 6.55, leading to a huge percentage error for 196.

In the following sections (Section 4.4.2.2 and 4.4.2.3), this study will only compare the prediction performances of different models evaluated using 2016-2019 data (i.e., the models are trained using 2015-2018 data, respectively). Year 2015 is excluded for model evaluations because the system had a very large-scale expansion at the early stage of system development and the model performances in 2015 are very different to the following years.

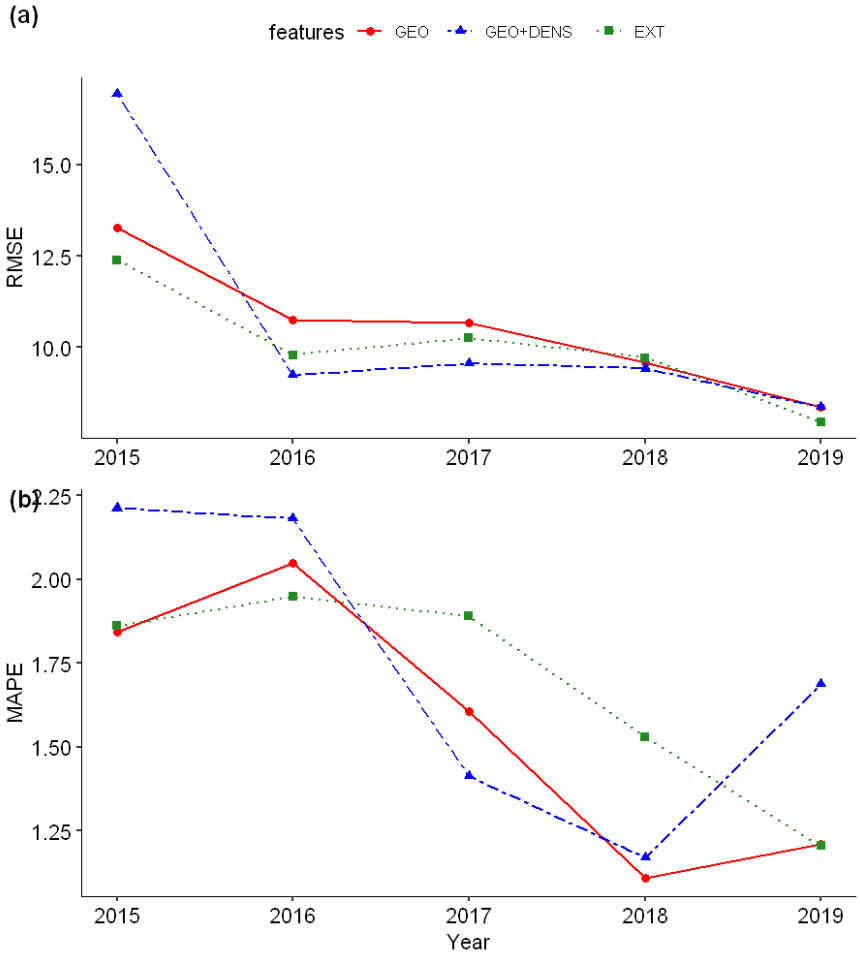


Figure 4.5. Model performance on all stations in different months: (a) RMSE, (b) MAPE

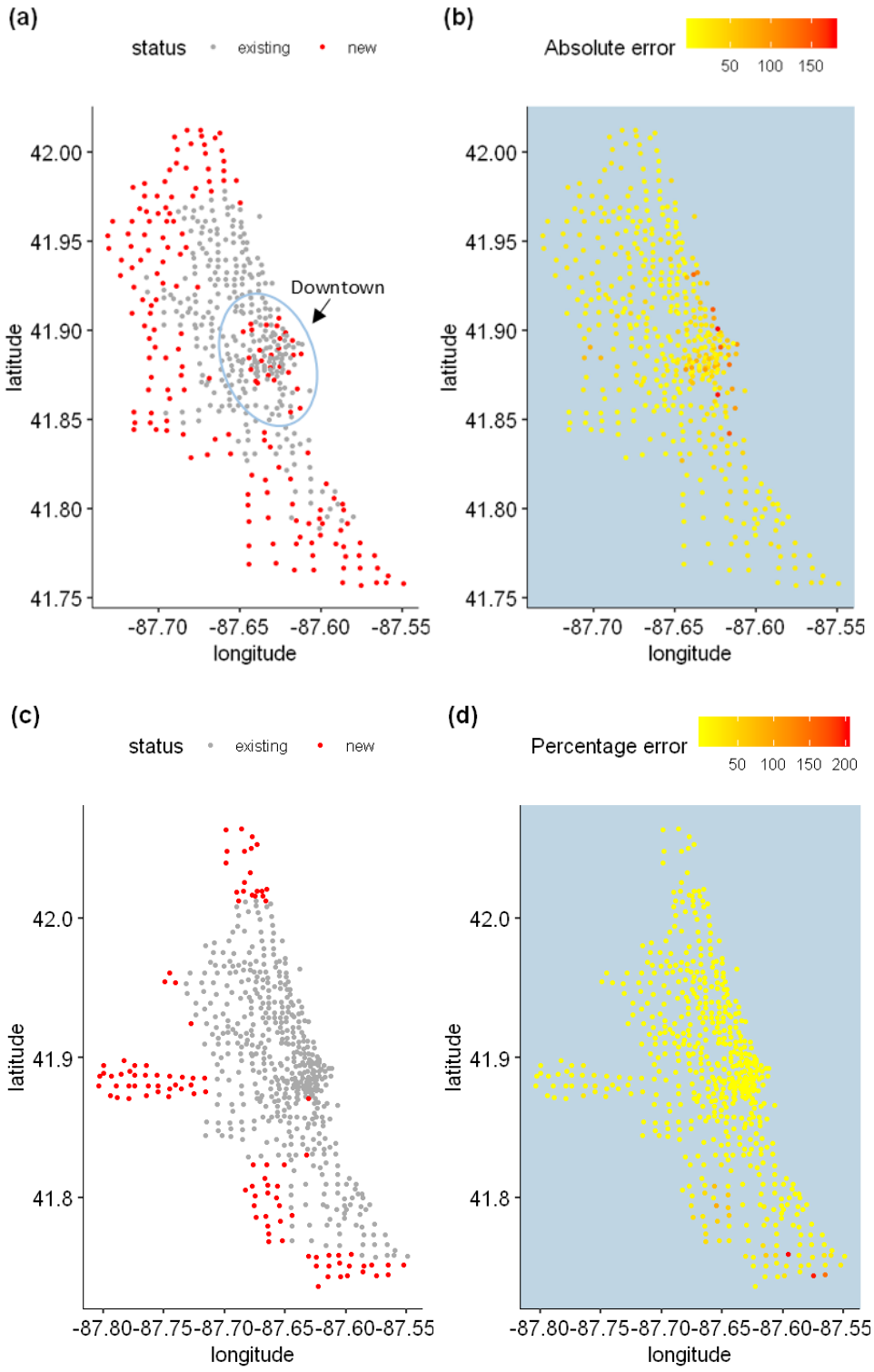


Figure 4.6. Spatial visualization of the prediction errors of "GEO-DENS" model in July 2015 ((a) locations of new stations, (b) absolute error) and August 2016 (((c) locations of new stations, (d) absolute percentage error)

After further exploration, it is found that the "GEO+DENS" model outperforms "GEO" model in the central area of the city by considering the changes of station density in the concentric buffers, but it tends to overestimate the demands in the outskirt areas, which leads to larger MAPE for these low-demand stations. Considering the whole system, the benefits from better predictions in the central high-demand areas outweigh the poor predictions in the outskirt low-demand areas. Therefore, the "GEO+DENS" model pays more attention to the spatial structures of the stations in the central areas, but such knowledge learned from the central areas does not apply to the outskirt areas. In contrast, the simpler model, "GEO", which only considers the latitude and longitude of stations, yields better predictions for those low-demand stations.

Figure 4.7 can help understand the difference between the predictions of the "GEO" and "GEO+DENS" models. For the high-demand stations in the red ellipse, the predictions from the "GEO+DENS" model are closer to the blue "perfect-prediction" line, compared with the predictions from the "GEO" model. In contrast, the "GEO" model performs better than the "GEO+DENS" model for those low-demand stations in the green ellipse. From the perspective of the whole system, by considering the changes of station densities, the reduced absolute errors by "GEO+DENS" for those high-demand stations reduced more of the overall absolute errors and could benefit more to the whole system. Take station 85 as one example, the "GEO" model predicts its average daily withdrawals to be 197 and the "GEO-DENS" model predicts it as 227; while the observed value is 287. Although both models under-estimate the demand for this station, the prediction from "GEO-DENS" is 30 trips closer to the observed value than the prediction from "GEO" model. Compare to the station network in July 2016, there were three new stations within 4.5 miles to station 85. The three stations were located in the bands of 0.7-0.8, 1.7-1.8, and 1.9-2.0 miles to station 85, which are in the distance range that stations complement each other. Therefore, the concentric station density features help the model understand the changes in the station network and improve the prediction performance. From the perspective of system planning and operations, reducing the prediction error by considering the network effects are more important for stations with high demand, such as station 85 which has more than 200 average bike withdrawals in a day. Having a more accurate prediction can help the system operators to better modify the station capacity, allocate bikes to stations, and estimate the total demands for a system expansion plan (obviously, demand changes at high-demand stations have larger influence to the total demand).

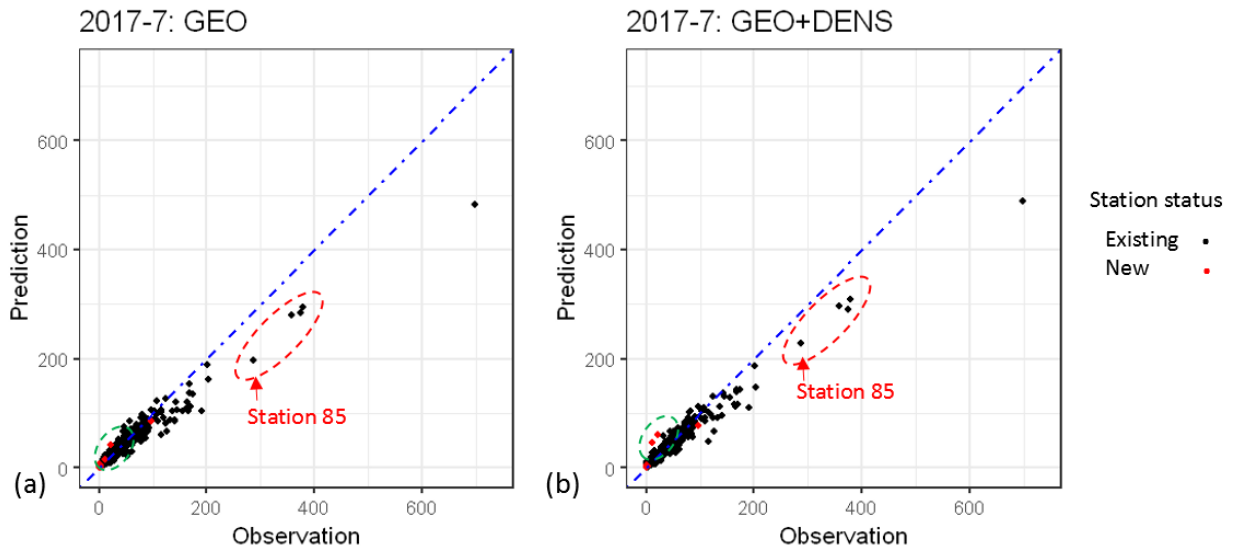


Figure 4.7. Target and predicted values for the predictions in July 2017 (the blue diagonal indicates perfect prediction): (a) the "GEO" model, (b) the "GEO+DENS" model.

4.4.2.2 Performance improvement by SEQEM

Based on the observations from the performances of "GEO" and "GEO+DENS" models, the proposed SEQEM is applied to further improve the prediction performances by setting a threshold of the spatial eccentricity and applying different models accordingly. Figure 4.8 shows the overall performance of SEQEM with different spatial eccentricity quantile threshold (evaluated using all the 2016-2019 data). When the spatial eccentricity quantile increases from 0.02, the RMSE decreases and reaches the lowest level when the quantile is between 0.7-0.8, where the RMSE is even lower than that of the "GEO+DENS" model. Also, the MAPE stays at a level that is close to the performance of the "GEO" model when the quantile is small, then reaches the lowest point when the quantile is 0.78. Considering both RMSE and MAPE, a quantile of 0.78 is identified as a quantile threshold that yields the best overall performance. This threshold can also be viewed as the spatial boundary only within which the station interactions should be considered. Figure 4.9 shows the changes of the spatial boundary corresponding to the 0.78 spatial eccentricity quantile threshold in June of 2014-2019. Demand predictions should consider simpler models (e.g., "GEO" model) for those outskirt stations identified by this boundary. Using the 0.78 quantile threshold, the SEQEM achieves a better performance than the "EXT" model regarding both RMSE and

MAPE. This indicates that the spatial structure of the station network contains very important information for the small-scale expansion demand prediction. Compared with "EXT" model, the SEQEM model has better transferability to be applied to other cities, which does not require external data.

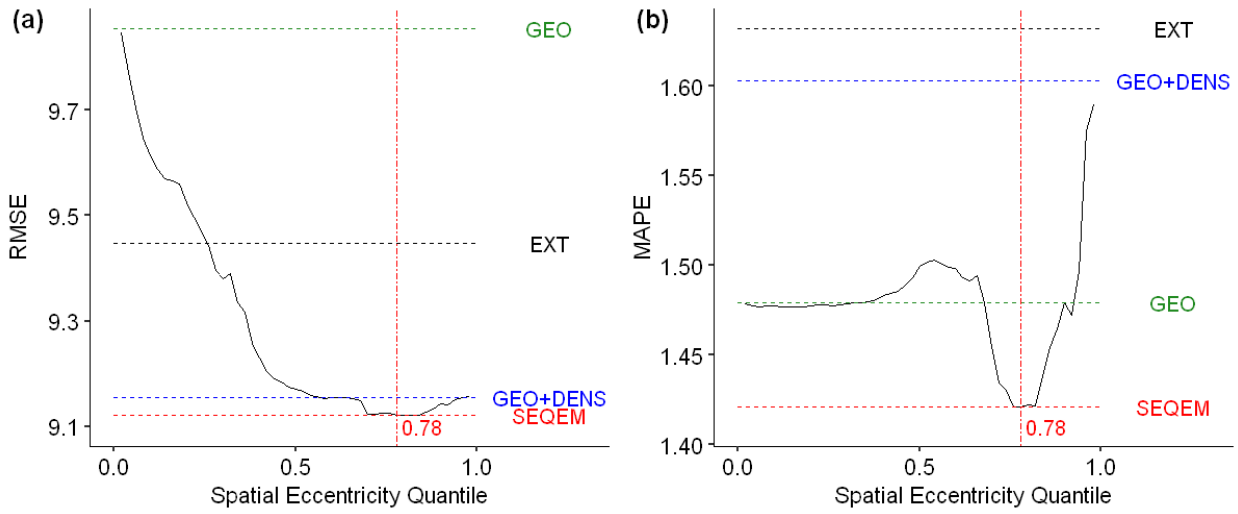


Figure 4.8. Performance of SEQEM by varying the spatial eccentricity quantile (the horizontal lines indicate the performance of the corresponding models in the right hand side), (a) RMSE and (b) and MAPE.

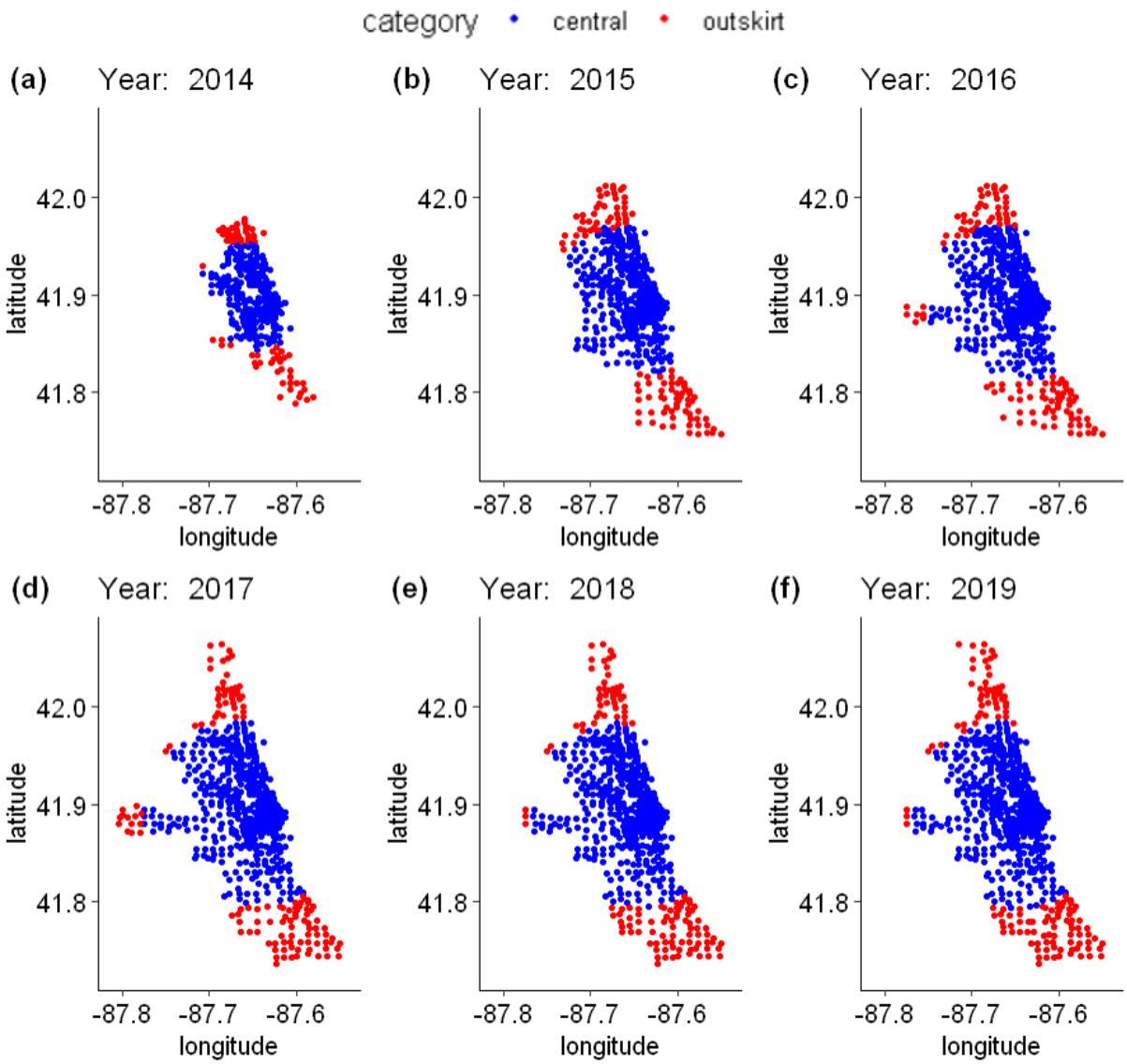


Figure 4.9. Changes of the spatial boundary corresponding to the 0.78 spatial eccentricity quantile threshold in June of 2014-2019

The SEQEM uses spatial eccentricity quantile as a threshold to combine the prediction results of the "GEO" and "GEO+DENS" models. However, there are other options such as simply adding the spatial eccentricity or spatial eccentricity quantile as a feature into the model (i.e., the "GEO+DENS+ECC" model and "GEO+DENS+ECCQ" model, respectively). These two models are also trained and evaluated. The results show that, for RMSE (Figure 4.10(a)) and MAPE (Figure 4.10(b)), both "GEO+DENS+ECC" (RMSE: 10.37, MAPE: 1.52) and "GEO+DENS+ECCQ" (RMSE: 9.22, MAPE: 1.54) yield poorer performance than SEQEM (RMSE: 9.12, MAPE: 1.42). Therefore, the spatial eccentricity quantile should only serve as a

threshold to identify a spatial boundary, but its value does not add more useful information for the prediction.

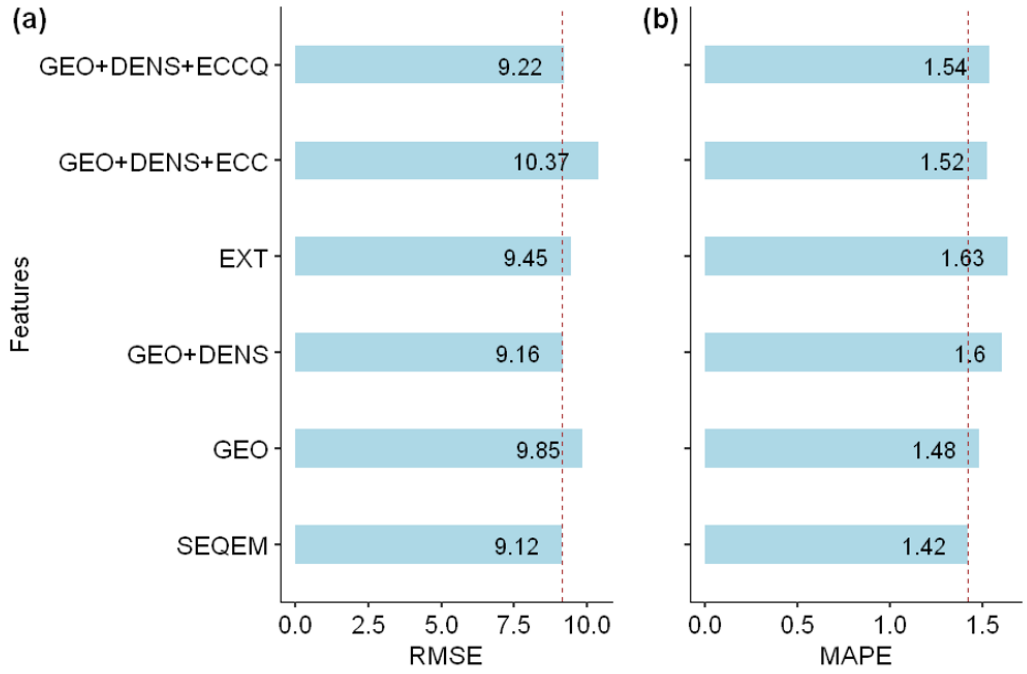


Figure 4.10. The performances when adding spatial eccentricity ("ECC") and spatial eccentricity quantile ("ECCQ") features: (a) RMSE, (b) MAPE (the vertical dashed lines indicate the performance of SEQEM)

4.4.2.3 The impact of concentric band construction

Besides the proposed more fine-grained "DENS" concentric bands to quantify the station network layout, another "DENS2" model is also implemented according to the previous study by Hyland et al. (2018), which only considers the number of stations within two distance bands: 0-0.5 miles and 0.5-3.1 miles. By comparing the performance of models that only uses "DENS" or "DENS2" features (Figure 4.11), it is found that the "DENS" model has a much lower RMSE of 10.31 than that of the "DENS2" model (22.97); "DENS" model also has a lower MAPE (2.47) than that of "DENS2" (2.99). Overall, the more fine-grained distance bands can better reflect the structure of the station network and lead to better prediction performances.

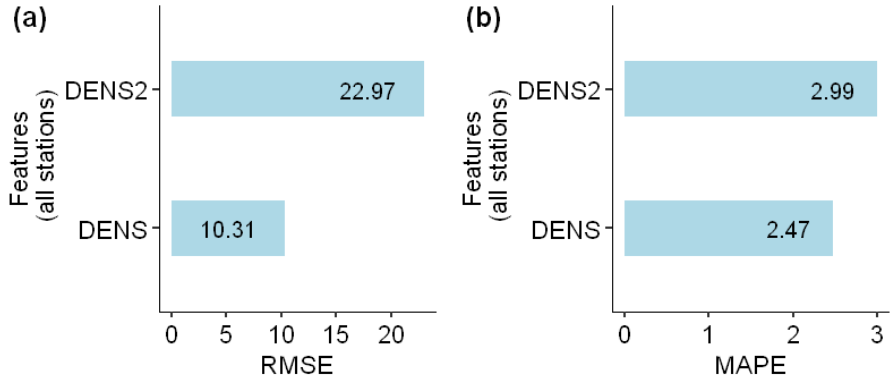


Figure 4.11. The performance on all stations using "DENS" and "DENS2" features: (a) RMSE, (b) MAPE

Lastly, by examining the feature importance of the "DENS" model, the distance ranges that are more important for predicting demand are analyzed. Note that one "DENS" model is trained for each month, thus the feature importance scores of different months' model may vary. Therefore, the average importance scores of different distance bands among different months' "DENS" model are computed and presented in Figure 4.12. The band distance that has the highest importance score is 0.25 miles (corresponding to 0.2 - 0.3 miles band) and followed by the 0.35 miles (corresponding to 0.3 - 0.4 miles band), which is close to the 0.30-mile change point identified by the piecewise regression model. Therefore, the feature importance analysis further strengthens the conclusions from the piecewise regression.

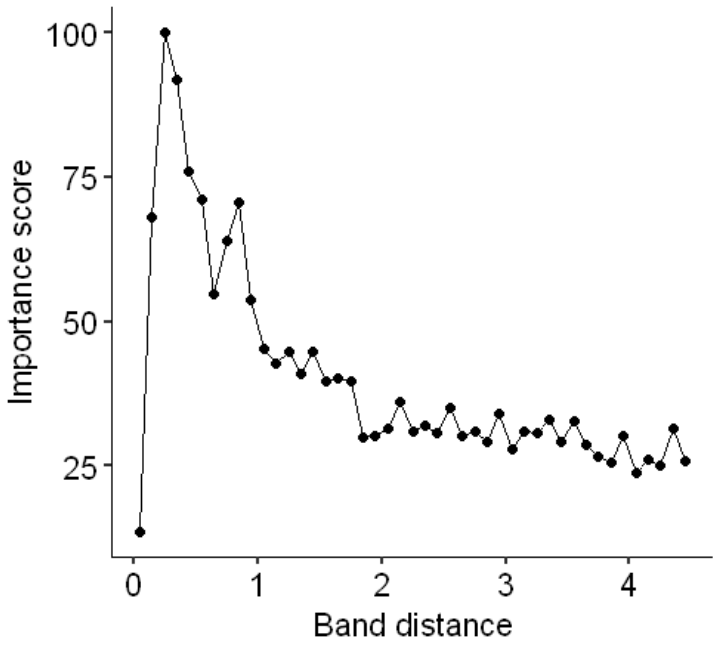


Figure 4.12. Average feature importance for different band distances

4.5 Conclusions and limitations

In this chapter, the station interactions in the BSS expansion process is analyzed using a piecewise regression model. According to the results of the piecewise regression analysis, the distances between stations contain very important information about station interactions and should be considered for expansion demand predictions. Then features of spatial station density in fine-grained concentric bands around a station are constructed to represent the number of stations in different distance ranges that the station can interact with. Such features are integrated into the demand prediction models to improve the prediction performances. A Spatial Eccentricity Quantile based Ensemble Model (SEQEM) is proposed to further improve the prediction performances and also identify the spatial range that the station interactions take effects.

The piecewise regression model shows that existing stations receive maximum complement benefits when the nearest new station is 0.3 miles away; but when the distance is less than 0.11 miles, the new station will dilute the demand of existing stations. This implies that the designers of a station-based BSSs should be more cautious of the cost-benefit tradeoff when placing stations too close (less than 0.11 miles). Also it would improve the siting of new stations to consider the

number of other stations that are around 0.3 miles away, which takes the most advantage of the station complement effects.

The results of the demand prediction models indicate that: integrating the station density in concentric distance bands improves the prediction performance by considering the spatial structure of the station network. "GEO+DENS" model performs well for central areas but has poor performance for outskirt low-demand stations. The proposed SEQEM addresses this limitation and improves the prediction performance. A 0.78 spatial eccentricity quantile is identified by the SEQEM as the threshold that yields the best prediction performance, which also indicates a spatial range within which the station interactions should be considered. With the 0.78 spatial eccentricity quantile threshold, the SEQEM yields better performance than the "EXT" model regarding both RMSE and MAPE. This indicates that the spatial structure of the station network contains very important information for the small-scale expansion demand prediction, which can also save the effort to collect and process the external data. The fine-grained distance bands also significantly improve the prediction performance than only using two bands in the previous study by Hyland et al. (2018). The feature importance scores from the trained "DENS" model reveal that the most important distance band for the prediction is the band of 0.2-0.3 miles, which is consistent with the change point at 0.3 miles identified from the piecewise regression model, which, in another aspect, supports the findings from the piecewise regression analysis.

This study provides unique insights for the station interactions in the BSS expansion process and practical suggestions to integrate spatial station network information to improve the BSS expansion demand prediction. However, there are some limitations that need to be improved in future research. First, the piecewise regression model only considers the impact from the nearest new stations, which assumes that the nearest new stations have the most impact to existing stations. Nevertheless, in the case that multiple new stations are added surrounding one existing station, the mixed effects from these new stations still require further investigations. Second, the piecewise regression model only analyzed the aggregated patterns of the station interactions. There could be other factors that affect the station interactions (e.g., locations of new stations, changes of local built environment, price promotions, etc.), which require more analyses in future studies. Third, the prediction results indicate that all models have poor performances for the large-scale expansion

in 2015, which should be improved in future research. Lastly, this study only focuses on the BSS in Chicago. Future studies can apply the piecewise regression model to other BSSs to analyze the differences in station interactions across different cities. Also, the SEQEM can be applied to other BSSs to evaluate the model transferability in different cities.

5. COMPARING THE PERFORMANCE OF DIFFERENT TYPES OF BIKE SHARE SYSTEMS

5.1 Introduction

As discussed in Chapter 1, there are three types of BSSs: station-based, dockless, and hybrid. This Chapter aims at comparing the performance of different types of BSSs. To assist the decision makings of selecting the system types that yield better user experience, lower operation cost, and less carbon emissions, this study proposed a stochastic simulation framework that reflects different bike parking requirements, spatial distribution of bikes and bike rebalance restrictions among different types of BSSs. The framework uses real-world bike share trip data to evaluate the performances of different types of BSSs from the perspectives of both user experience and operational challenges.

Different types of BSSs may affect user experience in terms of bike accessibility and the additional time required to pick up and return the bikes. Bike accessibility (how easy it is to find a bike) is important to users and thus impacts the ridership of the BSSs in the long-term. In station-based BSSs, the bikes are all located at the stations and where to find a bike is more predictable to the users. In the contrast, bikes are distributed more sparsely in dockless systems, making the locations of nearby bikes more uncertain. A hybrid BSS will be between the two, depending on the users' parking behaviors, i.e., whether more users park the bikes back to the stations or on the street. Compared to riding a personal bike that the users can directly travel from the trip origin to the destination, bike share users need to spend some excess time (Datner et al., 2015) in walking for bike pick-up/returning because the origin/destination of the bike trip (referred to as "biking O-D") may not be identical to the trip's actual origin/destination (referred to as "actual O-D"). Such excess time varies among different types of BSSs due to the different spatial distributions of bikes and the different bike parking requirements. In addition to walking to pick up a bike, bike return in station-based BSSs also requires extra time walking from the bike parking station to the actual destination. Furthermore, if the target parking station was full, the user would have to reroute to another station with available docks to return the bike, increasing the walking distance. Such bike returning excess time can be avoided in dockless and hybrid systems, since the users can directly

park the bikes at the final destinations. Long excess time can reduce user satisfaction of using bike share and cause the users to switch to other transportation modes.

Additionally, choosing different types of BSSs may require different bike rebalance and management considerations. The demand in a BSS is often imbalanced from both spatial and temporal perspectives. The system operators normally deploy vehicles to rebalance the bike fleet (Brinkmann et al., 2015; Fishman et al., 2014; Schuijbroek et al., 2017). In station-based systems, the number of bikes that can be allocated to a station is restricted by the station capacity (i.e., the number of docks). Certain number of docks may need to be left empty for bike return. In the case that a station has high bike-checkout needs that exceed the station capacity, some of the trips could be unserved unless more frequent rebalance is conducted, which causes higher operational costs. In the contrast, as there is no capacity limit in dockless or hybrid systems, more bikes can be allocated to high-demand areas. However, two new issues emerge: (1) because bikes are spatially more scattered instead of only being located at the stations, the rebalance vehicles need to travel around to collect the bikes; and (2) bikes parked on the sidewalks may crowd the street, obstruct the pedestrians, and cause safety concerns.

Although selecting the appropriate system type is important for developing new BSSs and expanding the existing ones, very few studies have quantitatively compared the above-mentioned differences in using different types of BSSs to guide the decision making. Most of the existing studies focus on improving the system performance of a specific type of BSS, but rarely consider the impact of changing system types (there are more discussions about relevant literature in Section 5.2). The few studies discussing different types of BSSs are either macro-scope analysis comparing the bike share use, policy, and business models across multiple cities through literature and government documents review (Gu et al., 2019; Hirsch et al., 2019; Susan et al., 2014) or analysis of the different usage patterns when both station-based and dockless BSSs are available in the same city (Chen et al., 2020a; Ji et al., 2020; Lazarus et al., 2020; Li et al., 2019). However, the differences observed from these studies are insufficient to support the decision making of which type of BSS works better to serve the demand in a specific city, because the observed differences may be the results of many external factors, such as population density, pricing schemes, service area assignment, and the support from local governments, etc. Therefore, to better guide BSS

development, there is a critical need for models that can compare different types of BSSs and quantify their system performances from the perspectives of both users and operators when serving the same demands.

To address this research gap, this study proposed a model that simulates the operation of different types of BSSs to serve the same demands and evaluates the system performances including service level (percentage of served trips), excess time spent by the users, vehicle-miles-traveled for rebalancing, and the number of bikes parked on street. The model uses bike share trip data as input to estimate the real-world demands, accounting for the unsatisfied demands that are not captured in the trip records. To evaluate how system types impact the spatial distribution of bikes and the bike flows, the model includes components to estimate the actual origins and destinations of trips and simulate user behaviors of bike pick-up and parking choices. To compare the operational needs in different types of BSSs, this study also integrated an optimization model into the simulation process for bike rebalance. The model is applied to three case study cities – Los Angeles, Philadelphia, and Chicago – to demonstrate how the simulation results of different types of BSSs can shed light on BSS improvement and guide potential system type changes.

The rest of this Chapter is organized as follows. Section 5.2 reviews relevant literature and highlights the contributions of this study. Section 5.3 presents the details of the proposed model and the input data of the case study cities. The results of the case studies are presented in Section 5.4. Section 5.5 concludes this Chapter and discusses potential future work.

5.2 Literature review

Most of the existing bike share literature focuses on station-based systems, because of their longer history. More data are also available for station-based systems, because many station-based BSSs publicly share their trip data. To guide the BSS operations, researchers have proposed models to predict the demands at different bike share stations (Hulot et al., 2018; Lin et al., 2018; Médard de Chardon & Caruso, 2015; Yang et al., 2016). Due to the spatiotemporally imbalanced demands, bike share operators normally use automobiles (e.g., vans) to move the bikes from the oversupplied docking stations to undersupplied ones. Without proper bike rebalance, bike share users may encounter empty/full stations, which would cause unsatisfied demand or increase the excess time

for the users to return the bikes. Because of the importance of rebalancing in BSS operations, many studies have proposed algorithms for optimizing bike rebalance (Alvarez-Valdes et al., 2016; Brinkmann et al., 2015; Caggiani & Ottomanelli, 2012; Chemla et al., 2013; Chiariotti et al., 2018; Fricker & Gast, 2016; Kloimüllner et al., 2014), aiming to improve the efficiency of the system by determining the number of bikes to be moved and the routes of the rebalancing vehicles. Overall, the abovementioned literature provides insights to improve existing station-based BSSs.

In recent years, because dockless BSSs' increasing popularity, a growing body of literature has studied dockless BSSs. Similar to the literature of station-based BSSs, the majority of the dockless BSS studies focus on travel pattern analysis (Chen et al., 2020a; Ji et al., 2020), demand prediction (Liu et al., 2018; Xu et al., 2018a), and optimization of bike rebalance (Caggiani et al., 2018; Luo et al., 2020; Zhang & Mi, 2018). In these studies, the bike share trip demands were aggregated to small regions (e.g., grids) to simplify the problem (Liu et al., 2018; Pan et al., 2019). Then the problems can be analyzed or solved at the region/grid level using similar methods as for a station-based system. These studies help improve dockless systems, but do not directly compare dockless BSSs to other types of BSSs.

A small number of studies evaluated the differences between station-based and dockless BSSs. While different types of BSSs have been launched in different cities, these BSSs often differ from each other regarding policies, business models, and pricing schemes. Gu et al. (2019) compared the station-based and dockless BSSs in China regarding market scales (e.g., the number of systems and bikes), government supervision, user demographic information, and financial sustainability. Similarly, Hirsch et al. (2019) summarized the dockless bike share literature and real-world practice in North America and pointed out that the business models of dockless BSSs are more profit-oriented than that of the existing station-based BSSs. Different from these studies that summarize the observations from multiple cities, some researchers analyzed the co-existence of station-based and dockless BSSs in the same city, where dockless BSSs are introduced in addition to existing station-based BSSs. Such co-existence of station-based and dockless BSSs is different from the hybrid BSSs, because the two systems are independently managed by different operators and the dockless bikes are not allowed to be parked at the docks of the station-based BSSs. Using trip data, researchers compared the ridership, user characteristics, and travel patterns between the

two types of BSSs. Li et al. (2019) analyzed the competition between the dockless BSSs and the station-based BSS in London and found that the average weekly usage of each station was reduced by 5.93% due to the newly added dockless BSSs from multiple companies. Lazarus et al. (2020) compared the Ford GoBike and JUMP (dockless e-bike share) in San Francisco. They found that dockless e-bike share trips have longer distances and durations than the docked trips. The JUMP system also has a higher bike usage intensity (daily trips per bike). Chen et al. (2020) found that the station-based BSS and the dockless BSS have similar user characteristics (in terms of gender, age, education background, etc.) based on a user-survey conducted in Hangzhou, China. Ji et al. (2020) compared the trip distances and durations between a station-based BSS and a dockless BSS in Nanjing, China. They concluded that, while the distance and riding time distributions are identical between the two systems, significant differences exist regarding the spatiotemporal usage patterns. While these studies identify how station-based and dockless BSSs perform differently or similarly among multiple cities or within the same city, the differences observed regarding ridership and the spatial-temporal usage patterns are also the results of many external factors, such as population density, pricing schemes, and the assignment of service areas. Such external factors are not tied to a specific type of BSSs (e.g., the same pricing scheme can be applied to both station-based and dockless BSSs). Therefore, results from the above studies cannot be used to conclude which type of BSSs can yield better performance.

Models that can compare the system performances of different types of BSSs to serve the same demand in a city are needed to guide BSS development. With the increasing focus of sustainable transportation globally, many cities are in the progress of planning their bike share service, while some other cities are considering switching their existing station-based BSSs to dockless or hybrid ones (Hirsch et al., 2019). Currently, there lack models to quantitatively compare different types of BSSs in a city to support such decisions. To address this gap, this study proposed a simulation model to evaluate the system performance of using different types of BSSs in the same city from the perspectives of both the users and the system operators.

Because different BSS types will result in different bike distributions and trip flows, which will impact the system's service quality (in terms of served trips and users' excess time, etc.), to enable comparison among different types of BSSs, it is critical for the model to (1) consider the true

demands of bike share, (2) estimate the trips' actual O-Ds, (3) consider the user behaviors when withdrawing/returning bikes to represent the complex interactions between the users and bikes/docks, and (4) have an integrated optimization model to account for rebalance activities.

First, while historical trip data of a BSS provides essential information on the need for bike share, it only records satisfied demands, not the true demands. Having a different BSS type may better serve the unsatisfied demands in certain regions while underserving the demands in others. Therefore, the comparison should be based on estimated true demands instead of observed historical demands. Previous studies (Caggiani & Ottomanelli, 2013; O'Mahony & Shmoys, 2015) have proposed methods for true demand estimation using historical data.

Second, models that can estimate the trips' actual O-Ds using the recorded biking O-Ds are needed to differentiate the user experience in different types of systems. Bike share trip data only records the biking O-Ds, i.e., where the users picked up and returned the bikes. When estimating demands for BSSs, existing studies often use the biking O-Ds (in most cases, bike share stations) as if they were the actual O-Ds (Alvarez-Valdes et al., 2016; Chen et al., 2018; Chiariotti et al., 2018; Datner et al., 2015; Freund et al., 2017; Huang et al., 2018), ignoring the walking stages and the excess time that the users had to spend. However, because different types of BSSs will have very different bike distributions which will significantly impact the bike pick-up and returning process, it is important to account for the walking stages and quantify the users' excess time in the comparison.

Third, better user behavior models are needed to account for the complex bike pick-up and return choices. Previous studies that model station-based BSSs often assume that a user's starting station is the actual origin of the trip and the user would leave the system unserved if the station is empty (Alvarez-Valdes et al., 2016; Chen et al., 2018; Freund et al., 2017). However, there may be other stations with bikes that are near the user's actual trip origin and the user may choose to reroute to pick up a bike from those stations. Some studies have considered such roaming behaviors between stations (Kabra et al., n.d.; Zhang et al., 2016c), but the users' walking behaviors are arbitrarily defined, assuming that the probability of walking to a station decays linearly with increasing distance, which may not represents the actual user behaviors in the real world. Similarly, the model should also be able to simulate the user behaviors to find a nearby bike in a dockless/hybrid system,

where the bikes are more scattered. Modeling the behaviors that users make the decisions on whether walking to pick up a bike or leaving the system is crucial for comparing the service levels among different system types. In addition, at the end of a trip, the positions that users park the bikes are influenced by the parking restrictions of the BSS types and, in the case of station-based BSSs, the dock availability. Tracking the bike parking locations is important to analyze the on-street-parking issues. The parking locations of one trip can also influence the subsequent trips, and thus affects the system performance of service level and excess time.

Lastly, although many studies have proposed rebalance methods for BSSs (Alvarez-Valdes et al., 2016; Brinkmann et al., 2015; Caggiani & Ottomanelli, 2013; O'Mahony & Shmoys, 2015; Schuijbroek et al., 2017), very few studies have integrated bike rebalance optimization models in simulations to quantitatively compare the rebalance needs among different types of BSSs.

To address the above discussed research gaps, this study proposed an integrated modeling framework that simulates the complex dynamics and the interactions among the users, bikes, stations, and the rebalance activities in different types of BSSs (station-based, dockless, and hybrid systems). The model also makes the following contributions in the context of existing BSS modeling literature: (1) this study proposed a method to infer the actual O-Ds of bike share demands, which enables better estimation of the excess time in the walking stages at the beginning and the end of a trip; (2) this work developed a user behavior model that simulates the walking behaviors according to the historical walking distance distribution when people uses walking as a transit mode, which better estimates the users' willingness to walk to access another transport mode; and (3) the model integrated a bike rebalance optimization model that accounts for the different rebalancing operations (i.e., station capacity restrictions in station-based BSSs and the routing to collect scattered bikes in the dockless/hybrid BSSs) in different types of BSSs.

5.3 Simulation model

5.3.1 Overview of the modeling framework

The simulation model mainly includes two steps: (1) demand estimation and (2) system simulation (Figure 5.1). The demand estimation step generates bike share demands with estimated actual trip

O-Ds. Using the historical bike share trip data from station-based BSSs and station status data (Section 5.3.2.1), the model first estimates the true bike withdrawal rates during different time periods at each station to address the demand censoring issue (Section 5.3.3.1). The biking trips are then stochastically generated based on the true bike withdrawal rates and the biking O-D correlations observed from the historical trip data (Section 5.3.3.2). After separating the service area into small grids, the model infers the trips' actual O-Ds based on the walking distance distribution from travel survey (Section 5.3.3.3). The generated travel demands (with start time and actual O-Ds information) are used as inputs in the system simulation step. The simulation model includes a user behavior model to determine how the travel demands are served and a bike rebalance model to periodically redistribute the bikes. The user behavior model simulates the user behaviors of bike pick-up based on the walking distance distribution and models the bike return in accordance with the operation rules of different system types (Section 5.3.4). Bike rebalance affects the availability of bikes (and docks in station-based BSSs). This study introduces an integer programing optimization model for station-based BSSs and a heuristic method for dockless and hybrid systems to redistribute the bikes (Section 5.3.5).

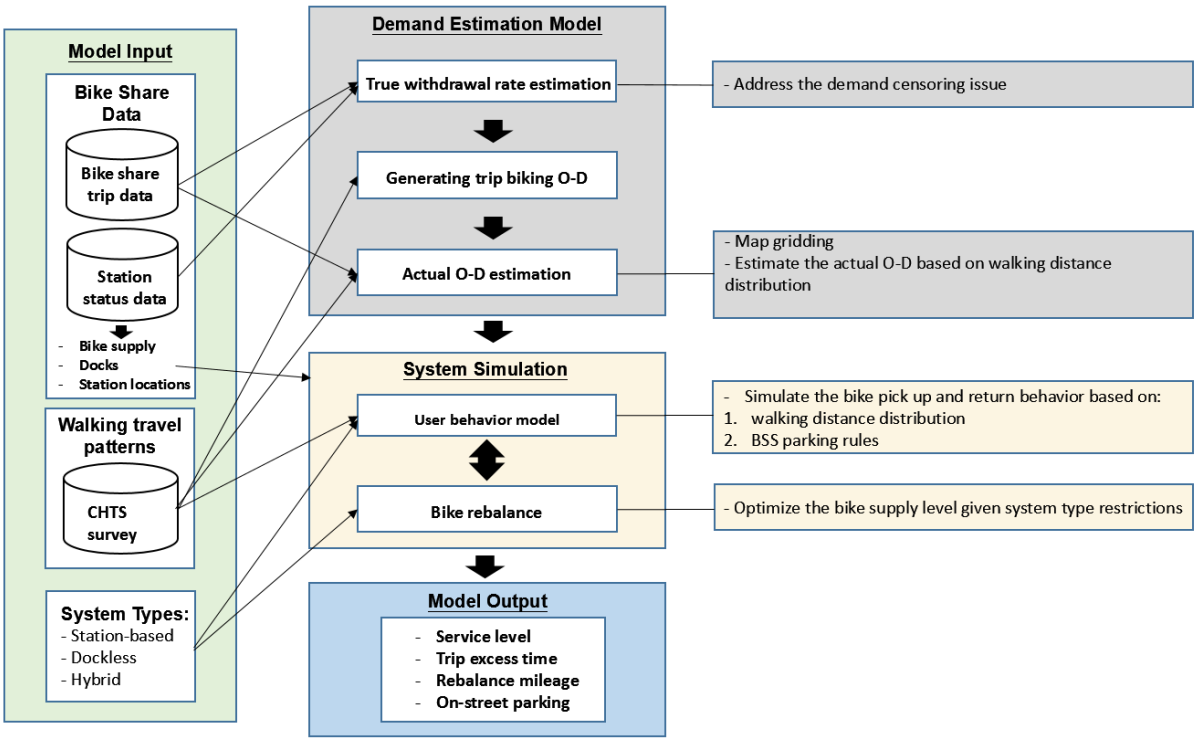


Figure 5.1. Overview of the modeling framework

Notation list:

B_k	The k^{th} (k = 1,2,3) grid band around the primary grid; the primary grid is represented as k = 1;
C_i	Capacity (the total number of docks) of station i;
C_i'	Working capacity of station i computed based the station's expected demand;
D_i	Auxiliary variables quantifying the expected imbalance of bike inventory at station i; a large value indicates that the station is either oversupplied or undersupplied;
e_k	Distance from the outer edge of the k^{th} band to the center of the primary grid;
f_τ^i	Accumulated net change of the available bikes at station i from the beginning of the planning horizon to the end of the τ^{th} hour ($\tau \in \{0,1,2,\ldots,Q\}$; a positive value means that station i is expected to have more bike returns than bike withdrawals;
g_i	The bike allocation weight for station i in a station-based BSS;
hp	Probability that the user is willing to end the trip at the nearest primary grid where docks are available instead of parking the bikes on the sidewalk in the actual destination grid (in hybrid BSSs);
L_i	The target inventory level (the number of bikes) of station i for bike rebalance;
M_t	The correlation matrix of biking O-D (primary-grid) pairs during hour t of the day, calculated from historical bike share trip data;
m_t^{ij}	The probability that a trip ends in station j given that it starts from station i in hour t of the day;
N	The total number of bikes (bike fleet) in the bike share system;
NB_k	The number of grids in B_k;
p_t^i	The percentage of time that station i had available bikes during hour t of the day;
Q	Rebalance interval; the rebalance activity is conducted every Q hours;
$RB_{k,\,first/last}$	Total probability of the users walking from all grids in B_k to a primary grid to pick up bikes ($RB_{k,\,first}$) or returning bikes to a primary grid and then walking to the actual destination in B_k ($RB_{k,\,last}$);
$RG_{k,\,first/last}$	The probability of the users walking from one grid in B_k to a primary grid to pick up bikes ($RG_{k,\,first}$) or returning bikes to a primary grid and then walking to one actual destination grid in B_k ($RG_{k,\,last}$); $RG_{k,\,first/last} = RB_{k,\,first/last} / NB_k$;
r_t^i	The average bike withdrawal rate during time interval t at bike share station i;

R_t^i	Estimated true average bike withdrawal rate during hour *t* of the day at bike share station *i*;
S	The set of stations in the station-based or hybrid BSS, indexed by *i* or *j*;
T	The number of days that the data is used for computing the average demand rate for all the stations;
u_i	Total expected bike checkout at station *i* during the planning period of bike rebalance;
v_i	Total expected bike return at station *i* during the planning period of bike rebalance;
w_{tz}^i	The number of observed bike withdrawals during hour *t* at station *i* on the z^{th} day;

5.3.2 Model Input

The proposed model takes three inputs: (1) bike share data, (2) walking distance distribution when people walk to transit to/from another mode, and (3) the type of system to be modeled.

5.3.2.1 Bike share data

The objective of this study is to compare the system performance of different types of BSSs when serving the same demands. Therefore, the most fundamental input is the bike share demands, which should reflect the temporal variations (e.g., hourly changes) of bike share demand in different regions in a BSS. The demands could either be learned from historical data (for cities that already have a BSS) or estimated from a demand prediction model (for cities that do not yet have a BSS). Obviously, historical data, if available, can provide a clearer picture about the demands than the predicted ones. In this study, the demand input is based on historical data from station-based BSSs because station-based BSSs' data are generally more available and some cities are considering transferring the existing station-based BSSs to dockless or hybrid ones (City of Philadelphia, 2019; The City of Chicago, 2019). But data from dockless/hybrid systems can also be used as model input (discussed in Section 5.3.3.4).

To estimate the true demands at existing stations (process explained in Section 5.3.3.1), the historical trip data is combined with station status data. The trip data contains the timestamp, station ID of the biking O-D, as well as the duration of each trip. Because the trip data only records historically satisfied trips, it may underestimate the bike withdrawal rate at a station, due to the issue of demand censoring (Goh et al., 2019; O'Mahony & Shmoys, 2015). This happens when

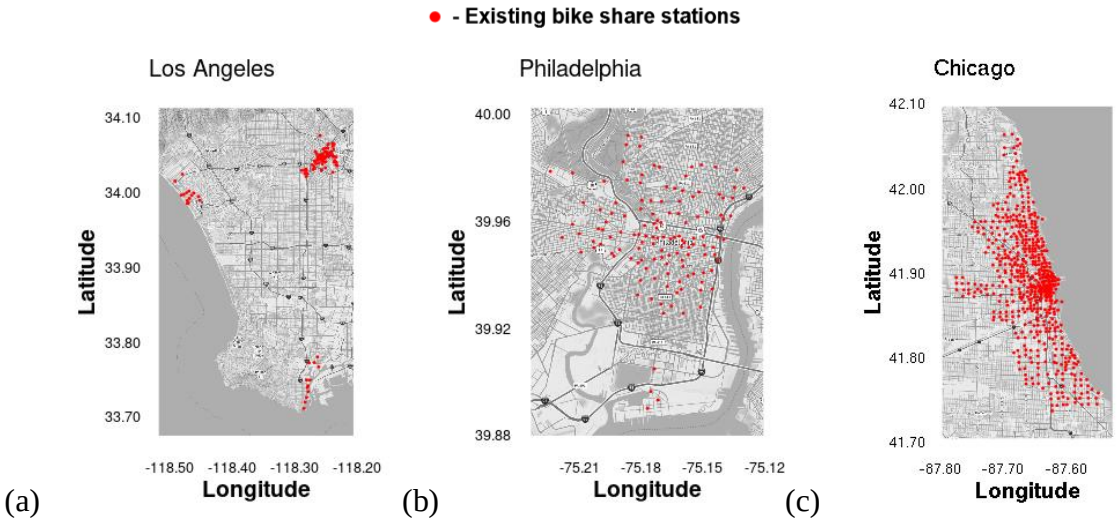
the station becomes empty or full, which prevents the users from checking out or returning bikes, respectively. To identify which stations were impacted by demand censoring during which time periods and better estimate the true demands, the station status data was collected at 10-minute intervals using the APIs provided on the BSSs' websites. The station status data contains the timestamp of data collection and the number of available bikes and docks at each station at the moment. Compared to other studies collecting such station data using 30-minute (Datner et al., 2015) or 15-minute (Schuijbroek et al., 2017) intervals, 10-minute intervals are chosen to gather more fine-grained information about station status.

For case studies, this study collected bike share trip and station data from three station-based systems: the Metro Bike Share (Metro Bike Share, 2019) in Los Angeles (including a small number of stations in Santa Monica and Long Beach), the Indego Bike Share (Indego Bike Share, 2019) in Philadelphia, and the Divvy Bike Share (Divvy Bike Share, 2019) in Chicago. Chicago's BSS is one of the largest BSSs in the United States regarding the daily trips; while the BSSs in Los Angeles and Philadelphia have relatively small and medium ridership, respectively (Table 5.1). This study selected these three BSSs with different sizes to study the performance of different types of systems under various usage intensities. The locations of the bike share stations in the three BSSs are presented in Figure 5.2. The case studies used data from the 20 weekdays in September, 2018. Previous studies have shown that bike share usage shows different temporal patterns between weekdays and weekends but very similar patterns exist in different weekdays (weekends) within a month (McKenzie, 2019; O'Brien et al., 2014; Zhou, 2015). This study focused on the weekdays in this analysis because they have more trips than the weekends. The method can also be applied to weekend data as well as data in other months and in other cities.

This study also counted the number of available bikes at 12 AM on September 17, 2018 (Monday) using the station status data as the input for current bike supplies in the simulation model (Table 5.1). The impact of different bike fleet sizes in the system is evaluated in the sensitivity analysis (Section 5.4.3.2). The number of docks at each station as well as the station locations are also extracted from the station status data. The number of docks will be used as the station capacity to model station-based BSSs, while the station locations provide information about the spatial distribution of bike share demands.

Table 5.1. Basic system size statistics of the analyzed BSSs in September 2018

City	Los Angeles	Philadelphia	Chicago
Average trips per day	986	2,387	15,404
Number of stations	93	131	581
Number of bikes	964	1,140	3,930
Number of docks	1,958	2,561	9,904



(a) (b) (c)

Figure 5.2. Existing bike share stations as of September 2018 in (a) Los Angeles, (b) Philadelphia, and (c) Chicago

5.3.2.2 Walking distance distribution

The proposed model uses the walking distance distribution from historical data to estimate the actual O-Ds (Section 5.3.3.3) and model user behaviors of the walking stage before picking up bikes (Section 5.3.4). This study extracted the trip data that involved walking from the 2010-12 California Household Travel Survey (CHTS) (California Department of Transportation, 2013). In the survey, trip information such as the travel distance and transportation mode was collected through wearable GPS devices so that the trip distances can be more accurately collected than the values reported by the survey takers. In the survey data, when a trip involved multiple travel modes, the trip is separated into different stages by mode. This study extracted the trips that had at least two stages and used walking as either the first (17,615 trips) or last (38,635 trips) stage's mode (e.g., walked to a bus stop and then took the bus). This data can help us understand how far people walked to connect to or from another transportation mode. In the following sections, the walking

trips that are in the first and last stages of those multi-stage trips are denoted as $WALK_{first}$ and $WALK_{last}$, respectively.

5.3.2.3 System type

For different system types (station-based, dockless, or hybrid), this study applys different bike parking requirements (Section 5.3.4) and bike rebalance (Section 5.3.5) in the model. For station-based BSSs, the bikes must be returned to stations; the bike rebalance is constrained by the station capacity and needs to consider leaving docks open for bike returns. In contrast, there are no such restrictions in dockless and hybrid systems.

5.3.3 Demand estimation

This section explains how the input data is used to estimate true demands and the trips' actual O-Ds.

5.3.3.1 True bike withdrawal rate estimation

Bike withdrawal from a bike share station is often modelled as a Poisson Process (Alvarez-Valdes et al., 2016; Chemla et al., 2013; Schuijbroek et al., 2017; Tao & Pender, 2017). The average number of bike withdrawals within a certain time interval (e.g., 1 hour in this case study) can be computed using Eq. (5.1).

$$r_t^i = \frac{1}{T}\sum_{z=1}^{T} w_{tz}^i \tag{5.1}$$

where r_t^i is the average hourly bike withdrawal rate during time interval t (e.g., 8:00 am to 8:59 am) over T days (determined by the estimation period, 20 days in this study) for a bike share station i and w_{tz}^i is the number of observed bike check-outs from the station i during time interval t on the z^{th} day. Similar to the existing literature (Lin et al., 2018; Yang et al., 2019, 2016; Zhang et al., 2016b), the trips are aggregated into hourly bins to capture each station's hourly demand variations in a day, assuming that the demand rate is the same within the hour.

To correct for the demand censoring when the stations are empty/full, this study estimates the true bike withdrawal rate R_t^i by adjusting the demand rate r_t^i with the percentage of time that the station has available bikes, adopting the method used by previous studies (Alvarez-Valdes et al., 2016; Goh et al., 2019; O'Mahony & Shmoys, 2015):

$$R_t^i = \frac{r_t^i}{p_t^i} \tag{5.2}$$

where p_t^i is the percentage of time that station *i* had available bikes during time *t* over *T* days. Because station status data is collected every 10 minutes in this study, this study computed the percentage of 10-minute-intervals that the station *i* had at least one bike during hour *t* over *T* days as an approximation of p_t^i.

5.3.3.2 Generating trips' biking O-Ds

Using the estimated true hourly withdrawal rate for each bike share station, the model can stochastically generate bike pick-up needs and trip time using the Poisson Process. The expected number of bike share trips originated from station *i* in hour *t* equals to the estimated true hourly bike withdrawal rate R_t^i, and the inter-trip time between two consecutive bike withdrawals follows an exponential distribution with the same rate parameter. Once the starting time and starting station of a trip are known, the next step is to assign a destination station for the trip. To do this, this study first computed the biking O-D correlation matrix M_t for station pairs for each hour *t* of the day using the historical data. In this matrix, the element m_t^{ij} denotes the probability that a trip ends in station *j* given that it starts from station *i* in hour *t* of the day, which is computed using the trip data in the *T* days and also corrected based on the bike and dock availability of the corresponding stations in the station status data. The details of computing m_t^{ij} are presented in Section D.1 of the APPENDIX D. Based on this biking O-D correlation, a trip that starts from station *i* in hour *t* may end at station *j* with probability m_t^{ij} and the trip destination can be simulated.

5.3.3.3 Actual origin and destination estimation

So far, trips with trip time and the biking O-Ds have been generated. This section further accounts for the walking stages and estimates the trips' actual O-Ds. To do this, this study first generalized the system's service areas beyond the existing stations by dividing the area into squared grids with an equal grid edge size of 400 feet (122 meters). This edge size is selected for two reasons: (1) in $WALK_{first}$, 81% of the walking trips are within 400 feet, indicating that users are willing to walk within a grid to pick up a bike; and (2) for the three studied cities, the grid is small enough that each grid contains at most one existing bike share station, allowing us to model the behaviors that users have to switch to a nearby station when the nearest station is empty or full at the grid-level. This resolution provides a balance between spatial granularity and computational intensity.

Next, this study estimated the actual O-D for each simulated trip based on the biking O-D and the walking distance distribution. The grids that contain existing bike share stations are defined as the "primary grids". Although the biking stage of the trip starts/ends from/at the primary grids, when consider the walking stages, the actual trip origin/destination may locate in the grids that surround the primary grids. This study assumed that the distance from the actual trip origin (destination) grid to the corresponding primary grid follows the same distribution as the walking distances in $WALK_{first}$ ($WALK_{last}$), which reflects how far people walked to connect to (from) another transportation mode. This study also assumed that no one is willing to walk more than two grids and defined the area within two grids as the "catchment" of the primary grid. Based on each surrounding grid's distance from the center of the primary grid (assuming the station is located at the center of the primary grid), the catchment can be separated into three bands (Figure 5.3): B_1 (brown color), B_2 (light orange), and B_3 (grey). B_1 is the primary grid. The distance from the outer edge of the catchment to the center of the primary grid is 1,000 feet (0.19 miles). According to the walking distance distribution from CHTS survey, 95.6% and 96.0% of the walking trips in $WALK_{first}$ and $WALK_{last}$, respectively, are within 1,000 feet. Therefore, the defined catchment can reasonably cover the majority of people's walking range.

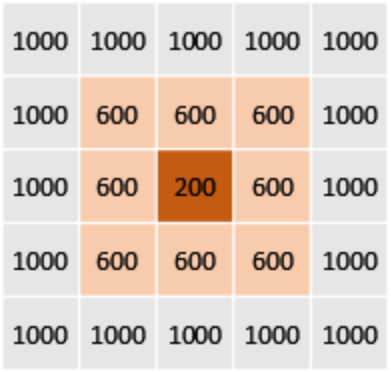


Figure 5.3. A primary grid (the center grid) and its catchment. The numbers in each grid shows the distance in feet between the outer edges of that band to the center of the primary grid)

Using the Empirical Cumulative Distribution Function (ECDF) of the walking distances from $WALK_{first}$, the total probabilities of the users walking from B_k (k = 1, 2, 3) to the center of the primary grid to pick up bikes can be calculated using Eq. (5.3).

$$RB_{k,\text{first}} = \begin{cases} ECDF_{\text{first}}(e_k) & k = 1 \\ ECDF_{\text{first}}(e_k) - ECDF_{\text{first}}(e_{k-1}) & k \in \{2,3\} \end{cases} \tag{5.3}$$

where e_k is the distance from the outer edge of B_k to the center of the primary grid. As a simplification, this study allocated the small portion of walking trips (4.36%) that are farther than 1,000 feet from the primary grid to B_3, ensuring that the sum of the probability in the catchment equals to 1. Assuming that all grids are equal within the band, the probability for each grid, $RG_{k,\text{first}}$, can be calculated using Eq. (5.4).

$$RG_{k,\text{first}} = \frac{RB_{k,\text{first}}}{NB_k} \tag{5.4}$$

where NB_k is the number of grids in the k^{th} band ($NB_1 = 1, NB_2 = 8, NB_3 = 16$). The calculated probabilities for each band and grid are listed in Table 5.2. Similarly, the probability of walking from the primary grid to each grid after bike return, $RG_{k,\text{last}}$, can be computed for destination estimation using the ECDF of the walking distance from $WALK_{last}$ (Eqs. 5.5-5.6).

$$RB_{k,\text{last}} = \begin{cases} ECDF_{\text{last}}(e_k) & k = 1 \\ ECDF_{\text{last}}(e_k) - ECDF_{\text{last}}(e_{k-1}) & k \in \{2,3\} \end{cases} \tag{5.5}$$

$$RG_{k,\text{last}} = \frac{RB_{k,\text{last}}}{NB_k} \tag{5.6}$$

Based on these calculated probabilities and a trip's biking O-D, the trip's actual O-D can be stochastically assigned to the grids. Take Chicago as an example, with 581 primary grids (581 existing stations), a total of 10,350 grids may serve as the actual trip origins or destinations. Figure 5.4 shows the expected number of trips per hour originating from each grid from 8 to 9 am in the downtown area of Chicago, illustrating how the demands have been redistributed to an expanded area surrounding the stations.

Table 5.2. Probabilities for assigning the grids in each of the three bands as trips' actual origins and destinations

Data	**Band B_k**	**e_k (feet)**	**ECDF(e_k)**	**Total probabilities for the band**	**Number of grids**	**Probability for each grid within the band**
$WALK_{first}$	B_1	200	52.12%	52.12%	1	52.12%
	B_2	600	89.68%	37.56%	8	4.70%
	B_3	1000	100%	10.32%	16	0.64%
$WALK_{last}$	B_1	200	55.65%	55.65%	1	55.65%
	B_2	600	89.69%	34.04%	8	4.26%
	B_3	1000	100%	10.31%	16	0.64%

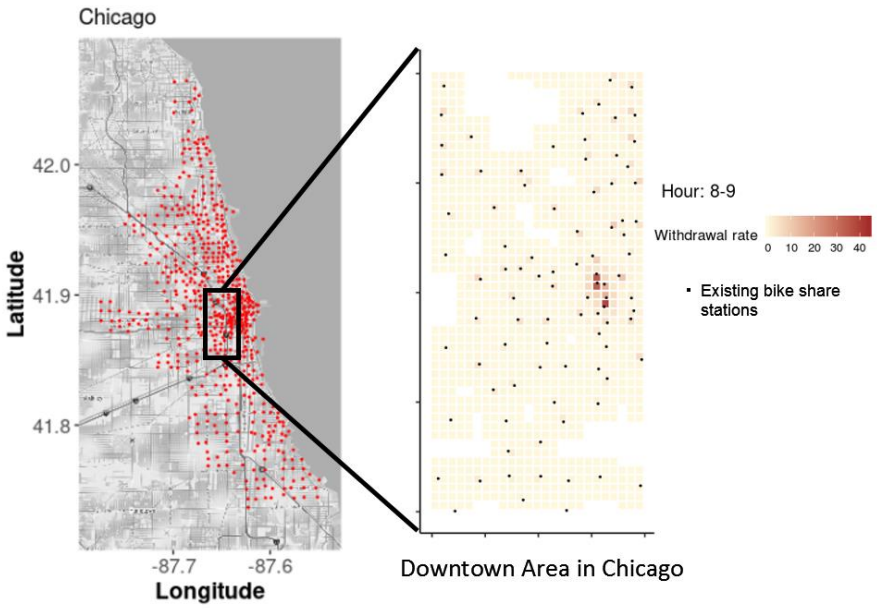


Figure 5.4. Distribution of the expected hourly bike withdrawal rate between 8 am and 9 am in Chicago downtown area

5.3.3.4 Demand estimation using dockless/hybrid BSS data

While this study used historical data from station-based BSSs as inputs, the above mentioned methods can also be applied to use data from dockless/hybrid BSSs as inputs with slight modifications. Trip data from dockless BSSs normally reports the latitude and longitude of the biking O-D of each trip. By aggregating the trips within the same grid, the same demand estimation process can be applied with two differences. First, to correct for demand censoring, the status of the dockless bikes need to be collected to identify the grids and the time periods that have no available bikes. Second, because dockless bikes can be parked very close to the actual destination, the biking destination can be considered as the actual trip destination.

5.3.4 User behavior model

Given the estimated demands with actual origin and destination grids for each trip, whether and how the trip will be served will be affected by the user behaviors, bike availability, and BSS types. When the grid of the trip's actual origin has available bikes, the user will randomly pick up a bike within the grid and start the trip. If this grid has no bike, the user will look for nearby bikes and make the decision on whether to pick up the nearest bike based on the distance he/she needs to

walk. This study assumes that the distance a user is willing to walk to pick up bikes follows the same distance distribution in $WALK_{first.}$ and is consistent with the revealed walking tolerance from the actual trip O-D assignment (more details explained in APPENDIX D, Section D.2).

After defining the above user behaviors in the walking stage of a trip, whether a trip will be served or not can be simulated. When a trip is served, the duration of the trip is computed using walking and biking distances and the average human walking speed of 276 feet/minute (3.13 miles/hour) (Levine & Norenzayan, 1999) and average biking speeds from historical data (6.23, 6.94, and 7.65 miles/hour in Los Angeles, Philadelphia, and Chicago, respectively). In the case that the bike is returned to the same grid (i.e., a round trip), the trip duration is assigned by randomly sampling the trip duration distribution from all historical round trips.

Besides the walking distance, another factor that affects how a trip is served is the bike parking rules of different types of BSSs. In a station-based system, the users are required to return bikes to the stations. So the users will have to look for the nearest primary grid around their final destinations, and then return the bikes to a random available dock at the station in the primary grid. In the case that there is no available dock at the moment in the primary grid, the users will need to reroute to the closest primary grid to the actual destination where docks are available. In a dockless BSS, the bike will be directly returned to the destination grid and parked on the sidewalk. In a hybrid system, because the users have the options to either park the bikes back to the stations or on the sidewalk, the parking behavior is more complex. The system operators may apply price incentives (Singla et al., 2015) to encourage users to park the bikes to stations, which would mitigate the on-street parking issues and also reduce the workload for bike rebalance when collecting bikes (since bikes are less scattered spatially). This study assumes that the users will return the bikes to the stations if his/her final destination is in a primary grid and there are available docks. Otherwise, with a certain probability *hp*, the user is willing to end the bike share trip at the nearest primary grid where docks are available instead of parking the bikes on the sidewalk in the actual destination grid. *hp* is a parameter that can be changed to represent different scenarios. This tests the impact of *hp* with values of 25%, 50%, and 75%. In reality, the probability *hp* can be affected by price incentives from the BSS, certain service rules, or other factors that could encourage the users to park bikes back to the stations.

5.3.5 Bike rebalance

Rebalance is an important part of the BSS operations. In station-based BSSs, rebalance helps avoid the situations of bike outage or full-docks, and thus can reduce the unsatisfied demands and improve user experience. One of the most important management decisions for bike rebalance is to identify the target bike inventory in a region to better serve the future demands. This study modified the method proposed by O'Mahony and Shmoys (2015) to allocate the existing bikes to different bike share stations. This resource allocation problem is expressed as an integer program in Eqs. (8) to (12). In a station-based system, the model allocates the total number of bikes N (the available bike supply in Table 5.1) in the system to each station (i.e., the primary grids), while the number of allocated bikes (inventory level) of station i is L_i and the dock capacity is C_i. Assuming that the system is rebalanced every Q hours, the goal is to find the initial inventory level for each station that best serve the demands in the planning horizon of the next Q hours. Based on the estimated true demand rate for each station, the accumulative net demand flow f_τ^i is computed, which tells us the expected net inventory change at station i from the beginning of the planning horizon to the end of the τ^{th} hour ($\tau \in \{0,1,2,\ldots,Q\}$, $f_\tau^i > 0$ when there are more bike returns than bike withdrawals at station i in the first τ hours of the planning horizon).

To more effectively allocate the bikes, this study introduced the working capacity C_i' for each station based on $\min_\tau f_\tau^i$ and $\max_\tau f_\tau^i$ (Eq. 5.7), which finds the capacity-in-use for each station based on the expected demand. Intuitively, a station i with $\min_\tau f_\tau^i < 0$ (having more bike withdrawals than returns) needs at least $|\min_\tau f_\tau^i|$ bikes at the beginning of the planning horizon to meet the consumption of bikes; while a station with $\max_\tau f_\tau^i > 0$ (having more bike returns than withdrawals) needs to leave at least $|\max_\tau f_\tau^i|$ docks open to serve the bike returns. Therefore, the working capacity C_i' is set to be $|\min_\tau f_\tau^i| + |\max_\tau f_\tau^i|$ for those stations that have $\min_\tau f_\tau^i < 0$ and $\max_\tau f_\tau^i > 0$ to guarantee that the expected bike withdrawals and returns can both be satisfied. On the other hand, when both $\min_\tau f_\tau^i$ and $\max_\tau f_\tau^i$ are positive/negative, the working capacity C_i' is

set to be $\max(|\min_{\tau} f_{\tau}^{i}|, |\max_{\tau} f_{\tau}^{i}|)$ to meet the maximum inventory change. Note that C_i' should be round up to the nearest integer and should not exceed the actual capacity C_i. The C_i in the model proposed by O'Mahony and Shmoys (2015) is replaced with the working capacity C_i' in constraint (10) to avoid allocating unnecessarily large number of bikes to stations with large capacity but low demand (more details are discussed in the Section D.3 of APPENDIX D).

The $L_i + \min_{\tau} f_{\tau}^{i}$ in constraint (5.10) denotes the expected minimum number of available bikes during the planning horizon, and correspondingly, the $C_i' - L_i - \max_{\tau} f_{\tau}^{i}$ in constrain (5.11) is the expected minimum available docks (according to the working capacity). In addition, to prioritize bike allocation to stations with high demand, this algorithm also introduces weights u_i and v_i in the constrains (5.10) and (5.11), respectively, where u_i is the total expected bike checkouts at grid i during the planning period and correspondingly, v_i is the total expected bike returns. The auxiliary variables D_i serve as a penalty for the bike inventory being too large or too small, with the consideration of the availability of both bikes and docks and station weights. Lastly, the objective function (5.8) aims to minimize the total penalty D_i of the entire system. Constraint (5.12) guarantees that L_i is non-negative and not greater than C_i'.

$$C_i' = \begin{cases} \min\{|\min_{\tau} f_{\tau}^{i}| + |\max_{\tau} f_{\tau}^{i}|, C_i\}, & \text{if } \min_{\tau} f_{\tau}^{i} \cdot \max_{\tau} f_{\tau}^{i} < 0 \\ \min\{\max(|\min_{\tau} f_{\tau}^{i}|, |\max_{\tau} f_{\tau}^{i}|), C_i\}, & \text{else} \end{cases} \tag{5.7}$$

$$\min \sum_i D_i \tag{5.8}$$

$$s.t. \quad \sum_i L_i \le N \tag{5.9}$$

$$\forall i \quad D_i \ge (L_i + \min_{\tau} f_{\tau}^{i}) u_i \tag{5.10}$$

$$\forall i \quad D_i \ge (C_i' - L_i - \max_{\tau} f_{\tau}^{i}) v_i \tag{5.11}$$

$$\forall i \quad L_i \in \{0,\ 1, ..., C_i'\} \tag{5.12}$$

To provide a fair comparison of different types of BSSs, this study sets different types of BSSs to have the same number of bikes, i.e., the current bike supply *N*. However, since the algorithm allows the summation of L_i to be less than *N* in constraint (5.9), it is possible to have unallocated bikes after solving the above resource allocation problem. This study does not directly allocate all the bikes in constraint (5.9), because the problem formulation may sacrifice bike return needs to allocate the additional bikes. Instead, when $\sum_i L_i < N$, the following Algorithm 1 is proposed to allocate the remaining bikes to stations with extra empty docks and low expected bike return v_i. In Section D.3 of the APPENDIX D, it is shown that the proposed method significantly improved the service level (percentage of unserved trips) compared with the original method by O'Mahony and Shmoys (2015), which does not consider the working capacity and the weights u_i and v_i.

Algorithm 1: Allocating remaining bikes in a station-based BSS

Input: The solved inventory level L_i from the integer program; the expected bike return v_i for each station *i*; Total number of bikes *N*

Output: Final target inventory level L_i for station *i*

Procedure AssignRemainingBikes(L_i , v_i, *N*)

WHILE $\sum_i L_i < N$:

$N_{remain} = N - \sum_i L_i$ # The remaining bikes to be distributed

$g_i = (C_i - L_i) \cdot (\max_i v_i - v_i)$ # compute the bike allocation weights g_i for station *i*; stations that have more empty docks and smaller expected bike return v_i will have a larger g_i

$L_i = \min(L_i + N_{remain} \cdot \frac{g_i}{\sum_i g_i}, C_i)$ # L_i round to integer

---

The rebalance of station-based system aims to balance bike return and withdrawal needs. Because dockless and hybrid BSS are not restricted by the station capacities, the problem formulation cannot be directly applied. Additionally, rebalance each grid in dockless and hybrid BSSs may be unnecessary when the total number of bikes in nearby grids are at a good level, since the grids in the same catchment can complement each other. Therefore, this study proposed a heuristic method to allocate bikes in dockless or hybrid systems at the catchment-level (Algorithm 2). Intuitively, when there's no capacity limit, bikes should be first allocated to those catchments with net bike

outflows (i.e., $\min_{\tau} f_{\tau}^{i} < 0$). Algorithm 2 prioritizes the primary grid catchments that have net bike outflows and uses u_i (total expected bike checkout from station i) as the weighting factor to assign the remaining bikes. When moving bikes into a catchment, the rebalanced bikes are only allocated to the primary grid instead of to each individual grid, because it is easier for the BSS operators to rebalance the bikes to demand hubs than to drop smaller amount of bikes to every grid.

Algorithm 2: Assigning Inventory Level for Primary Grid Catchments in Dockless/Hybrid Systems

---

Input: $\min_{\tau} f_{\tau}^{i}$ and u_i for each primary grid i;
Total number of bikes N
Output: Target inventory level L_i for each primary grid catchment
Procedure AssignInventoryLevelNoDocks($\min_{\tau} f_{\tau}^{i}$, u_i, N)

IF $-\sum_{i}(\min_{\tau} f_{\tau}^{i} \mid \min_{\tau} f_{\tau}^{i} < 0) \le N$: # sufficient fleet size to place the maximum outflow number of bikes in all grids

FOR each primary grid i:

IF $\min_{\tau} f_{\tau}^{i} < 0$: # First assign the bikes to grids that have net outflow

$L_i = -\min_{\tau} f_{\tau}^{i}$

ELSE $L_i = 0$

$N_{remain} = N + \sum_{i}(\min_{\tau} f_{\tau}^{i} \mid \min_{\tau} f_{\tau}^{i} < 0)$ # The remaining bikes

FOR each primary grid i:

$L_i = L_i + N_{remain} \cdot \frac{u_i}{\sum_i u_i}$ # Allocate the remaining bikes using u_i as the weight, L_i round to integer

ELSE: # the system has insufficient bikes to place the maximum outflow number of bikes in all grids

FOR each primary grid i:

IF $\min_{\tau} f_{\tau}^{i} < 0$: # First assign the bikes to grids that have net outflow

$L_i = N \cdot \frac{-\min_{\tau} f_{\tau}^{i}}{\sum_{i}(-\min_{\tau} f_{\tau}^{i} \mid \min_{\tau} f_{\tau}^{i} < 0)}$ # L_i round to integer

ELSE $L_i = 0$

After obtaining the list of grids that need rebalance, this study computes the mileage to travel through these grids by solving a traveling salesman problem as an estimation of the rebalance work load (more details in Section D.4 of the APPENDIX D). In dockless and hybrid systems, to

account for the vehicle travel needs to collect the sparsely parked bikes within a grid, bike locations are randomly assigned within the grid and the rebalance vehicle will need to make multiple "stops" to collect these bikes in the rebalance routing.

5.3.6 Model initialization, simulation, and output

The simulation model is initialized by allocating the bikes to the grids according to the assigned inventory level determined by the rebalance algorithms. Then the bike share use is simulated throughout the day using the demands generated from the demand estimation model as inputs. In the simulation process, the generated demands are loaded at their start time; the users search for nearby bikes, ride a bike (or leave the system), and return the bike according to the user behavior model. The rebalance activity will take place according to the rebalance interval. Because the information of how frequently the existing station-based BSSs perform rebalance is unavailable, to better simulate the existing station-based systems, this study evaluates how different rebalancing frequencies (e.g., every 2, 4, 8, 12, or 24 hours) impact system service (the number of trips being served) and identifies the rebalance frequency that yields the percentage of served trips (out of the estimated true demand) that is the closest to the historical data as the rebalance frequency in the model. To evaluate the impact of demand and user behavior uncertainties, for each scenario, 10 rounds of simulations are conducted using 10 sets of trips estimated using the same true demand rates but different randomness seeds. To ensure comparability across different BSS types, in each round of simulations, the loaded trips are exactly the same for different types of BSSs.

The model outputs provide system performance information from both the users' and operators' perspectives. For each served trip, the excess time is recorded for different stages: (1) start-walk: the excess time spent by users walking from the actual trip origins to the locations of a nearby bike; (2) end-walk: the excess time spent by users walking to their final trip destinations after parking the bike at a station; and (3) end-bike: the excess time spent by users riding the bike to nearby stations when the first-choice stations were full. In a dockless system, only the start-walk stage may generate excess time, since the users can always park the bikes directly at the final destinations. Besides excess time, the model also tracks the system's service level (the percentage of trips being served out of the total estimated true demands), the total mileage of rebalance vehicles, the

percentage of trips that parked bikes on the sidewalk, and the maximum number of bikes parked on the sidewalk of each grid (for dockless and hybrid systems) in the simulated day.

5.4 Results and discussions

This section first discusses the selection of rebalance intervals in Section 5.4.1 and then compares the different types of BSSs from the aspects of service rate, users' excess time, rebalance vehicle mileages, and bikes parked on the sidewalk in Section 5.4.2. In Section 5.4.3, the results of sensitivity analysis on adding Geo-fencing zones to dockless BSSs and also the impact of changing the bike supplies are discussed.

5.4.1 The selection of rebalance frequency

This study uses the service level of historical data as the benchmarks for choosing the rebalance frequency. The average service level is calculated as dividing the average number of daily trips in the historical data by the average number of simulated true demands per day (average of the 10 sets of trips). As shown in Figure 5.5 (a), the benchmark service levels in Los Angeles, Philadelphia, and Chicago are 98.4%, 93.2% and 87.0%, respectively. A very high service level in Los Angeles means that, in most of the time, the stations have available bikes and thus most of the trips can be served.

Next the station-based system is simulated over one day using different rebalance frequencies (rebalance every 2, 4, 8, 12, 24 hours or having no rebalance) to study the influence of rebalance. The results are presented in Figure 5.5 (b). "None" means a scenario that there's no rebalance, and the simulation started with the observed bike distribution from station data at 12 AM of September 17, 2018. For other rebalance scenarios, the bikes are first rebalanced at the beginning of the day based on the rebalance algorithm; and the rebalance is carried out at the specific intervals. As expected, the no rebalance scenario yielded more unserved trips than the benchmarks (horizontal lines) and more frequent rebalance leads to better performance. Even rebalancing every 24 hours yields less unserved trips than the benchmark. Therefore, in the following analysis, this study uses

24-hour (i.e., overnight rebalance) as the rebalance interval because it yields the closet performance to the historical data.

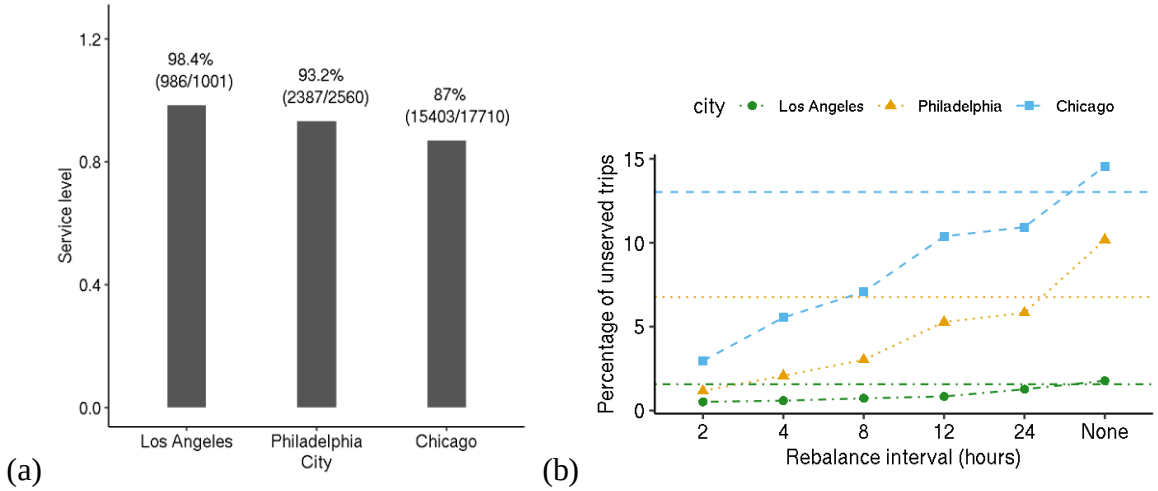


Figure 5.5. The service level of historical data and simulated station-based systems using different rebalance frequencies: (a) Service level of historical data according to the estimated true demand (the numbers in the parenthesis indicate "average number of observed daily trips/simulated true demands"), (b) Percentage of unserved trips in one day using different rebalance intervals (the horizontal lines show the percentages of unserved trips in historical data).

5.4.2 Comparing different BSS types

This section compares the simulation results of different types of bike share systems from both the operators and users perspectives, evaluating the system's service level, users' excess time, rebalance mileage, and on-street parking issues. For each city, the simulations were carried out with the same sets of generated trips, bike supply, and rebalance frequency (every 24-hour). The only difference is the system type.

5.4.2.1 Service level

Because users may leave the system unserved (e.g., choose another transportation mode or do not make the trip) when they cannot find available bikes within the regions that they are willing to walk to, having less unserved trips shows better system service. Figure 5.6 shows the percentage of unserved trips when using different types of BSSs in the three cities. In Los Angeles, only a very small portion (less than 1.5%) of the trips are unserved and the different system types do not

have significant impact on the system's service level. In Philadelphia, the dockless system significantly reduced the unserved trips compared to the station-based BSS, while in Chicago, the unserved trips in a dockless system are in a similar level as that in a station-based BSS. In both Philadelphia and Chicago, the hybrid system performed better than the station-based and dockless BSSs.

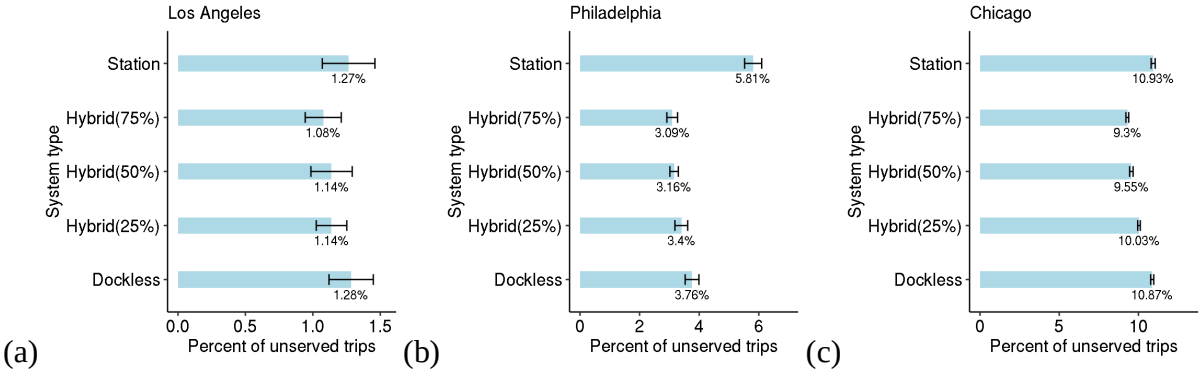


Figure 5.6. Percentage of unserved trips for different system types (for hybrid systems, the percentage shown in the parenthesis is the park-to-station probability)

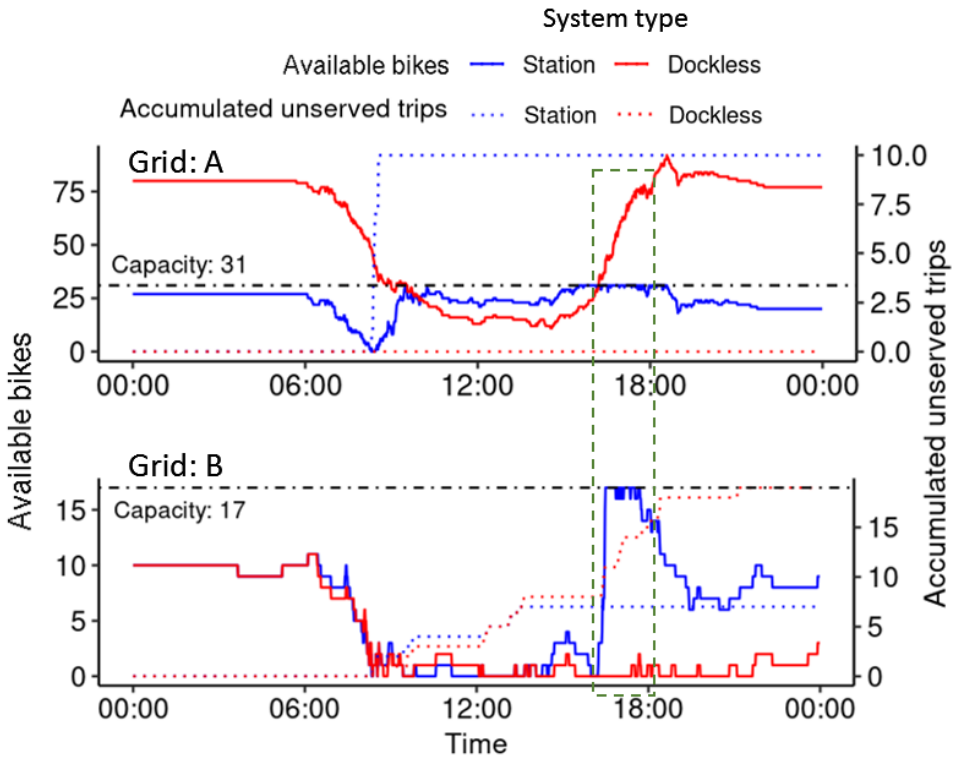
The distinctions of unserved trips between different system types are mainly due to three factors: (1) the bike allocations in different types of BSSs, (2) the scattered bike distributions in dockless/hybrid BSSs, and (3) the interactions of nearby grids. For bike allocation, changing the system type mainly change the maximum number of bikes that can be allocated to a grid. Station-based BSS, being constrained by the station capacity, has to balance between available bikes and available docks. On the other hand, in dockless and hybrid systems, it is possible to stock the grids with high bike-withdrawal demands with as many bikes as needed (if the system has sufficient bikes). Figure 5.7 (a) shows an example of how bike allocation and the change of inventory levels are impacted by the system types in a grid with high morning bike-withdrawal demand (Grid A) in Chicago. In the station-based system, although the station in grid A was filled with bikes at near-capacity level, at around 8:00, the station became empty for a short period and caused 10 unserved trips. In contrast, in the dockless BSS, since there is no capacity limit, more bikes can be allocated to grid A at the beginning of the day to avoid bike stock-out and better serve the anticipated high morning bike-withdrawal demand.

However, the flexibility of dockless and hybrid BSSs may cause the bikes to be more distributed and cause unserved trips in the primary grids. The Grid B in Figure 5.7(a) shows an example that, when having the dockless system, the bike inventory level in this grid is low because five users parked the bikes to other grids within the catchment of Grid B, instead of returning the bikes to Grid B if it was a station-based system. As a result, when a user in Grid B who later needed a bike but found no available bike in the grid, the trip could be unserved if the user is unwilling to walk to nearby grids to pick up a bike. In this model, the primary grids (where the stations are located in existing station-based BSSs) have higher demands than surrounding non-primary grids, which can be viewed as the demand "hotspots". When siting the bike share stations in a station-based BSS, the decision makers tend to place stations near demand hotspots such as commercial areas, subway stations, and living zones with high population density (García-Palomares et al., 2012; Park & Sohn, 2017). Therefore, dispersing the bikes to other locations could reduce the service in these demand hotspots. For hybrid BSSs, because the station structure are kept and some users may be incentivized to return bikes to the primary grids near their actual destinations, the bike availability at these demand hotspots can be improved, reducing the unserved trips than the corresponding dockless BSSs.

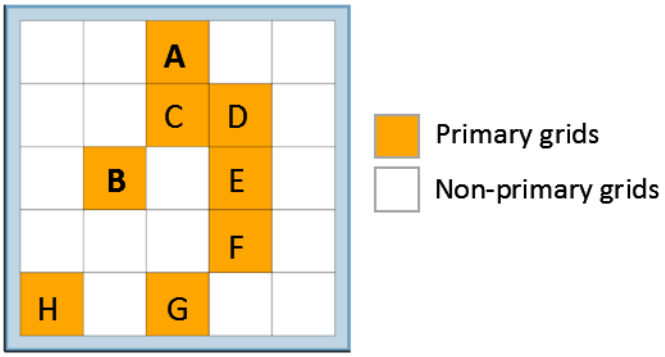
Additionally, the rerouting behaviors of users between neighboring grids also affect the bike availabilities and thus the service level. Take Grid B in Figure 5.7(a) as an example, it is interesting that, during the afternoon rush hour (in the green dashed box in Figure 5.7(a), there is an inventory level "jump" in the station-based system, which is not observed in the corresponding dockless BSS. The reason is that, during this period in the station-based BSS, some other primary grids (Grids A, C, D, E, and F as shown in Figure 5.7(b)) near Grid B are full, which forces the users to return bikes to Grids B and G where docks are available. These bikes were later used to serve the withdrawal demand in Grid B in the evening. However, since the dockless system has more flexible bike return requirements, such rerouting did not happen and the bike inventory remained low at Grid B after the afternoon rush hour, which led to more unserved trips in the remaining of the day than the station-based system. This example shows that, the forced rerouting due to capacity limits in station-based BSSs could indirectly help redistribute the bikes and improve service level in certain regions. This result also shows the need for future rebalancing algorithm to take into account the interactions of neighboring grids. For instance, Grid B had a $\min_{\tau} f_{\tau}^{i}$ of -10

in the 24-hour rebalance interval and thus the initial inventory was set to be 10 bikes, which would be sufficient to serve Grid B's demand if there were no interactions between nearby primary grid catchments. However, the simulation results showed that 12 bikes were picked up from Grid B to serve trips originated from nearby primary grids which encountered bike stock-out, which significantly changed Grid B's bike availability. Simulation-based optimization could potentially be used to better rebalance the systems but is beyond the scope of this study.

BSSs that have denser station networks are more likely to have such interactions between neighboring grids. This is reflected in the differences between the unserved trips in Philadelphia and Chicago. Changing the system type in Philadelphia made a much more significant impact than in Chicago because the denser station network in Chicago benefited more from the interactions between neighboring stations. In the station-based BSS in Chicago, in 16% of the trips (out of total true demands), the users rerouted to another station to pick up bikes, while only 6% of the trips rerouted in the station-based BSS in Philadelphia. This is because that in Chicago, more stations are closely located to neighboring stations. There are 64 and 174 primary grids located within the catchments of other primary grids in Philadelphia, and Chicago, respectively. In the downtown areas of the two cities, stations are distributed denser in Chicago than in Philadelphia. If taking the average for the first 100 shortest distances between nearest station pairs, the values are 0.22 miles in Philadelphia and 0.14 miles in Chicago. The majority of bike share demand is also concentrated in Chicago downtown (heat map of daily bike withdrawal rate of the stations is plotted in Figure D. 3in the APPENDIX D). In systems with denser station networks like Chicago, the benefit gained from removing the station capacity constraint in dockless/hybrid systems could be cancelled by losing the self-distribution of bikes through forced rerouting in station-based systems.



(a)



(b)

Figure 5.7. Example of the dynamic changes of two grids (A and B) in one round of simulation in Chicago: (a) temporal change of bike inventory level and the number of accumulated unserved trips, (b) relative locations of the two grids and other nearby primary grids

5.4.2.2 Excess time

Although forced rerouting helped to redistribute the bikes in a station-based BSS, the cost is users' excess time. As expected, the station-based systems have higher average excess time per served trip (Figure 5.8, the detailed results can be found in Table D. 1 in the APPENDIX D). The majority of the excess time (57 – 68%) in a station-based system is from the end-walk stage. This is because that, on average, the users spent more time walking to the final destination from a nearby dock at

the end of a trip than walking to a bike to start the trip. The full-dock issue that forced the users to reroute to return bikes (end-bike stage) accounts for 8-18% of the total excess time in the three cities. There were 3%, 6%, and 16% of the simulated trips encountered full-dock issues in Los Angeles, Philadelphia, and Chicago, respectively. Therefore, although the rerouting behavior at the end of trips in station-based BSS may help redistribute the bikes and thus reduce the total unserved trips, it actually sacrifices the user experience to a relatively large extent (e.g., 16% of the trips were affected in Chicago). Additionally, rerouted trips due to full-dock issues would also cause more excess time in the end-walk stage. Out of the total excess time in end-walk stage, 38%, 30%, and 57% of the excess time were incurred by the full-dock issues in Los Angeles, Philadelphia, and Chicago, respectively. In dockless and hybrid systems where the bikes could be distributed to a broader area, the average excess time in the start-walk stage is also shorter than that in the station-based systems. In the three analyzed cities, the users spent less time accessing bikes in dockless BSS than in station-based systems. The average excess time per served trip in the start-walk stage is about 10-second shorter in a dockless system (36-37 seconds) than in a station-based system (46-47 seconds). Overall, dockless BSSs saved 1.2 – 1.8 minute (Los Angeles: 1.5, Philadelphia: 1.2, Chicago 1.8) excess time per served trip compared with station-based BSSs. The average trip durations in Los Angeles, Philadelphia, and Chicago are 15, 12, and 12 minutes, respectively (historical September 2018 bike share trip data), so the saved excess time is about 10-15% of the actual trip travel time. Considering that bike share trips are relatively short-distance trips, such improvement of user experience is significant. To actualize such benefits, the bikes should have devices to accurately track their locations, and thus the users can locate the bikes using a smartphone application. The total excess time of the hybrid systems is longer than that of the dockless system and increases with higher park-to-station probability. However, removing the need for the users to reroute to other stations still have significant time-saving benefits.

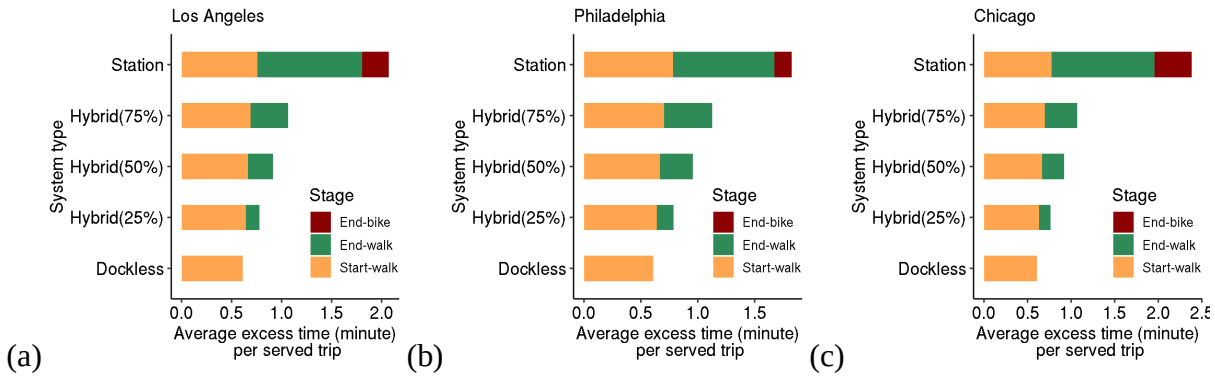


Figure 5.8. Average excess time (minute) per served trips

5.4.2.3 Rebalance mileage

Although the rebalance vehicles in a dockless/hybrid system will need to make more stops to collect scattered bikes, removing the station capacity may help reduce the number of primary grids that need to be rebalanced. Therefore, the net impact of system types on rebalance needs should also be evaluated (Figure 5.9 (a)). In Los Angeles, the dockless system requires slightly more rebalance mileages than the station-based BSS, although the differences across different BSS system types are small. In Philadelphia and Chicago, the average daily rebalance mileages in dockless systems is 10.5% and 10.8% more than that in the station-based systems, respectively. In Philadelphia, the rebalance mileages in the hybrid BSSs are similar to that in the station-based BSS; while in Chicago, the total rebalance mileage in hybrid BSSs is reduced by 4.1% to 8.1% compared to the station-based BSS. Similar to the unserved trips, the differences of the rebalance mileages are also caused by three factors: the different bike allocations, additional mileages required to collect scattered bikes in dockless/hybrid BSSs, and the interactions of neighboring grids which may affect whether a primary grid catchment requires rebalance.

The rebalance mileage differences between dockless and hybrid BSSs are mainly due to additional mileages required to collect scattered bikes. Depending on how scattered the bikes are, dockless BSSs require 4%, 16%, and 19% more rebalance mileages than the corresponding hybrid BSSs with a 50% park-to-station probability. In hybrid BSSs, when a primary grid has oversupplied bikes that need to be removed, the rebalance vehicle often only needs to visit the station to take out the planned number of bikes. However, in a dockless systems, the rebalance vehicle will have many more stops to collect the scattered bikes. Figure 5.9 (b) and (c) show that, in the hybrid BSSs,

a relatively small number of extra stops are required in Los Angeles (13 stops, subtracting Figure 5.9 (c) values from Figure 5.9 (b) values) and Philadelphia (3-10 stops), and Chicago (76-168 stops) to collect scattered bikes. In contrast, the corresponding dockless BSSs require 127, 214, and 995 extra stops in Los Angeles, Philadelphia, and Chicago.

The initial bike allocations and the interactions of neighboring grids could also affect which primary grid catchments need to be rebalanced, resulting different rebalance mileages when these catchments cover significantly different areas. In Los Angeles and Philadelphia, the visited primary grid catchments cover very similar spatial areas across different types of BSSs (Figure D. 4 in the APPENDIX D), so the total rebalance mileages in station-based BSSs are similar to that in the hybrid BSSs. However, in Chicago, the rebalance of the station-based BSS had to include some outskirt grids that are not visited in the corresponding dockless/hybrid BSSs, significantly increasing the total rebalance mileages (more discussions in Section D.7 of the APPENDIX D). Besides bike allocation, it is also observed that the interactions of neighboring grids may also affect whether a grid needs to be rebalanced. For example, due to receiving the bikes rerouted from nearby grids, Grid B in the station-based BSS in Figure 5.7 (a) had almost the same inventory at the end of the day (9 bikes) with the initial inventory (10 bikes), and did not require a rebalance visit. In contrast, the end-of-day inventory at Grid B in the dockless BSS is much further from the initial inventory, and thus needed to be rebalanced.

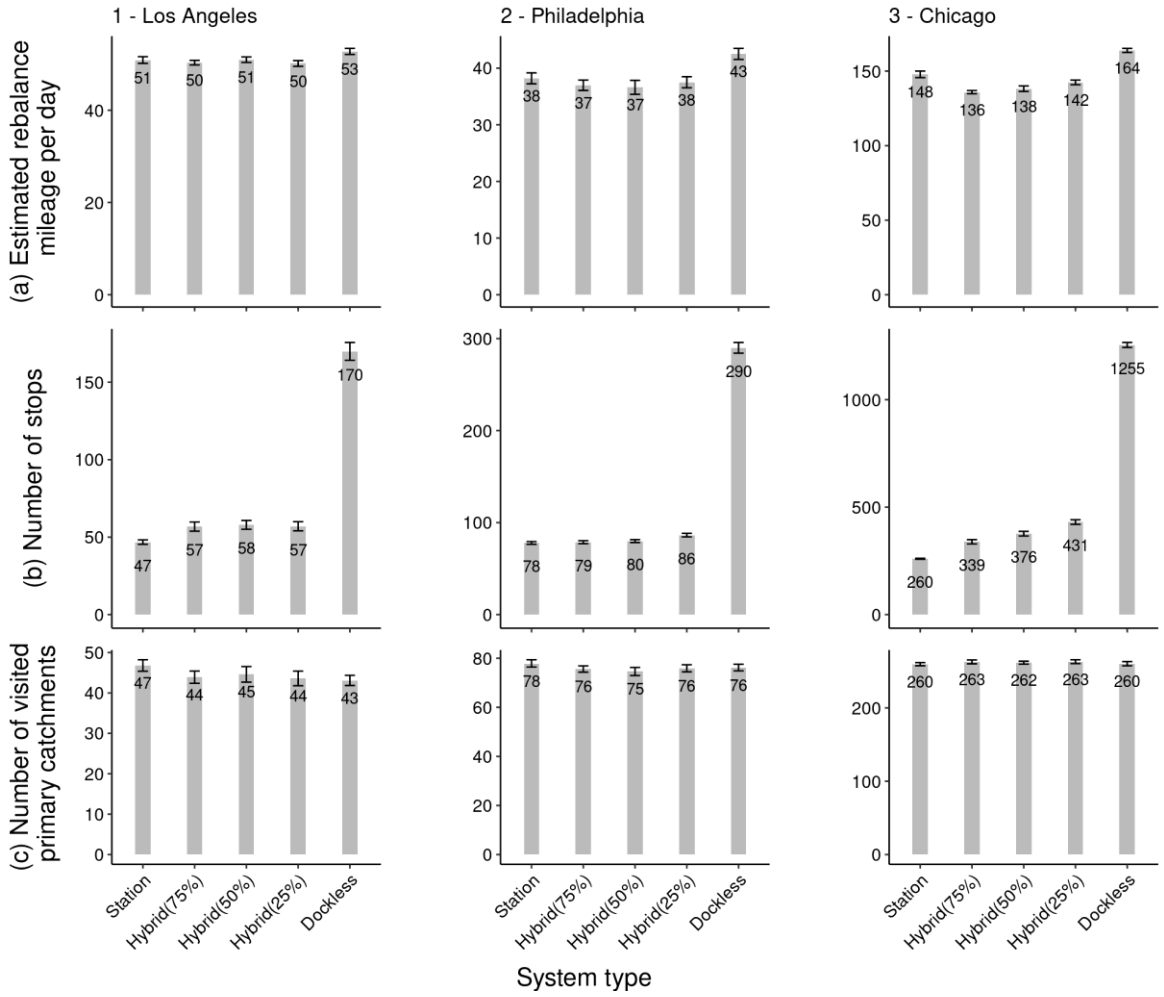


Figure 5.9. Rebalance workload in different types of systems: (a) estimated total rebalance mileages per day, (b) average number of stops the rebalancing vehicle has to make, and (c) the number of visited primary grid catchments

5.4.2.4 Bikes parked on the sidewalk

Clogging the sidewalks is a major concern for dockless bikes. In the simulations of the hybrid BSSs, this study assumes that the users are willing to return the bikes to the stations if their actual destination grids have stations; and when the actual destination grids do not have stations but the nearest station has available docks, with a certain park-to-station probability, the users are willing to reroute to the nearest station to return bikes. Figure 5.10 shows that, in hybrid systems, bikes are parked to stations in more than half of the trips. With 25% park-to-station probability, in 42%, 39%, and 38% of the served trips, bikes are parked on the sidewalks in Los Angeles, Philadelphia, and Chicago, respectively. The percentage of on-street parking trips decreases near linearly with a

higher park-to-station probability. It is notable that, in a hybrid system, a user only considers taking the extra effort to return the bike to the nearest station when there are available docks. To evaluate how full stations impact bike parking behaviors, this study simulates the scenarios with a park-to-station probability of 100% and finds the percentage of on-street parking trips are 22.3%, 15.9%, and 18.0% in Los Angeles, Philadelphia, and Chicago, respectively. These results show that, even when all users are willing to park the bikes to the nearest stations, the on-street-parking issues will still exist in the hybrid systems due to docks being full.

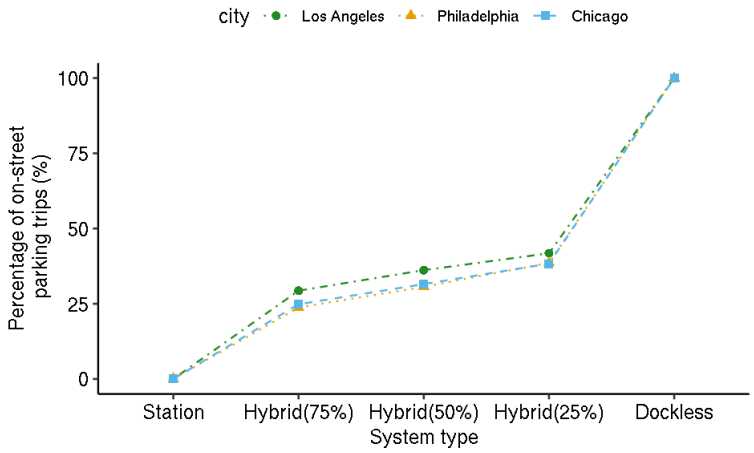


Figure 5.10. Percentage of served trips that bikes are parked on the street

For each simulated day, the maximum number of bikes that were parked on the sidewalk is also recorded for the dockless systems to evaluate the worst-case scenario that one grid may encounter. On average, the worst-case grid could have 64, 59, and 195 on-street bikes in Los Angeles, Philadelphia, and Chicago, respectively. In order to better understand how these bikes may crowd the grid, the grid with 195 bikes is plotted in Chicago (Figure 5.11). Each bike is randomly located within the grid, which covers an area of about one street block. The bikes are densely located and the average distance between any two nearest bikes is about 14.5 feet (4.4 meters). Given that bikes cannot be parked in buildings, the spatial distribution of these bikes would be even denser if all bikes are on the sidewalk. To better regulate dockless bike parking, some cities are implementing geo-fencing technologies to require users to park their bikes to designated parking areas on the sidewalk (Cheng et al., 2019; Zhang et al., 2019), which helps reduce the turmoil from undesired bike parking behaviors. However, this would add additional cost by adding radio frequency identification technology (RFID) devices to the bikes. In this example, if all of the 195

bikes were parked in a row at a geo-fenced area, this would require an area with a length of 244 feet (74 meters), assuming the average width that a bike takes is 15 inches (38 cm). This is feasible considering that the edge length of a grid is 400 feet and the geo-fenced zones can be located in both sides of the streets. The impact of adding geo-fenced zones with limited capacity in a dockless BSS will be discussed in Section 5.4.3.1.

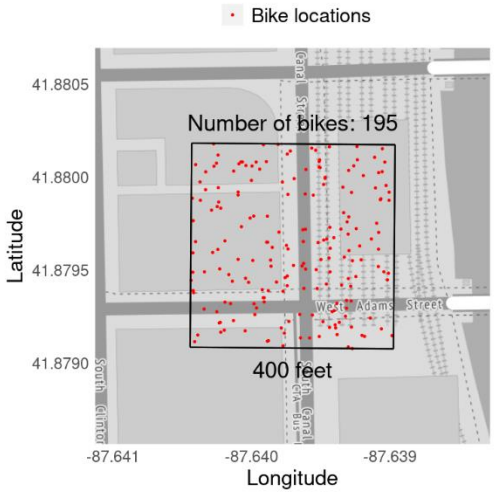


Figure 5.11. Distributing 195 bikes randomly in one grid in Chicago

5.4.2.5 Overall performance of the different system types

Overall, if ranking the three types of BSS according to the four metrics analyzed above and use the rank as a score (the best = 3, middle = 2, and the worst = 1), the hybrid system has the best overall performance when equal weights are assigned to the four metrics (Table 5.3). In Los Angeles where the bike share usage intensity is low, the hybrid system has almost identical overall performance with the station-based system. There are no significant improvements of service level and rebalance workload when switching from the station-based BSS to a hybrid BSS, while the on-street parking issues may offset the benefits of excess time for such transitions. Therefore, it may not worth to change the current station-based system in Los Angeles. The difference is more significant in Philadelphia and Chicago, where the BSS has much higher usage intensity. In these two cities, only if on-street parking outweighs the other aspects, the station-based systems would be the better choice. The benefit of hybrid systems regarding service level is more significant in Philadelphia than Chicago. As discussed in Section 5.4.2.1, a station-based BSS benefits the most

regarding service level from switching to a hybrid system when the demand is high but the existing stations are sparse.

Table 5.3. Performance summary of the three types of systems in Los Angeles

City	System Type	Service level	Excess time	Rebalance workload	Parking on the street	Overall score
Los Angeles	Station-based	Best (3)	Worst (1)	Best (3)	Best (3)	10
	Hybrid	Best (3)	Middle (2)	Best (3)	Middle (2)	10
	Dockless	Best (3)	Best (3)	Middle (2)	Worst (1)	9
Philadelphia	Station-based	Worst (1)	Worst (1)	Best (3)	Best (3)	8
	Hybrid	Best (3)	Middle (2)	Best (3)	Middle (2)	10
	Dockless	Middle (2)	Best (3)	Worst (2)	Worst (1)	8
Chicago	Station-based	Middle (2)	Worst (1)	Middle (2)	Best (3)	8
	Hybrid	Best (3)	Middle (2)	Best (3)	Middle (2)	10
	Dockless	Middle (2)	Best (3)	Worst (1)	Worst (1)	7

5.4.3 Sensitivity Analysis

In Section 5.4.2.4, this study discusses the potential of remitting the on-street-parking issues in dockless BSSs through adding Geo-fencing zones. Therefore, Section 5.4.3.1 analyzed how Geo-fencing zones with different zone capacities may impact the results. In addition, the above simulation results are based on the current bike supply identified in historical data. Because the bike supply may also affect the performance of the BSSs, Section 5.4.3.2 discusses the impacts from having different number of bikes in the systems.

5.4.3.1 The impact of Geo-fencing zones

When implementing geo-fencing zones, the zone capacity mainly serves as a limit when allocating bikes for rebalance. That is to say, the number of bikes allocated to a grid cannot exceed the designed zone capacity. For example, if the rebalance algorithm sets the target inventory level of a primary grid catchment to be 100 bikes while the zone capacity is only 40, instead of allocating

all 100 bikes to the primary grid, only 40 bikes would be allocated to the primary grid while the remaining bikes would be randomly redistributed to nearby grids without exceeding the zone capacity of those grids (first consider grids that are one-grid away, and then two-grid away). However, the zone capacity may be relaxed for bike return and allow the users to park bikes at a grid even if its inventory level exceeds the zone capacity, because the geo-fencing zones are more flexible than the docks in a station-based BSS and the users may squeeze in more bikes than the designed capacity. The number of grids exceeding its capacity during any periods in the day are also tracked. For bike rebalance, the rebalance vehicle only needs to stop once at the geo-fencing zone when visiting a grid, instead of stopping multiple times to collect scattered bikes.

Figure 5.12 shows the simulation results when altering the parking zone capacity. The unserved trips do not change with the zone capacity in Los Angeles and Philadelphia (Figure 5.12 (a)). For Chicago, smaller parking zone capacity helps reduce unserved trips in Chicago, although the absolute change is small (10.9% to 10.6%) when switching from purely dockless to dockless with a 30-bike zone capacity. For the three factors that affect the service level, adding the Geo-fencing zones does not change the bike allocations at primary grid catchment level. Only a small number of grids that need a large number of bikes and thus need to be adjusted according to the zone capacity. In the scenario of having a zone capacity of 30 bikes in the dockless BSSs, only six, two, and 13 primary grids have an initial inventory level that exceeds the capacity in Los Angeles, Philadelphia, and Chicago, respectively. Therefore, adding the zones only makes the bikes more broadly distributed in these small number of grids at the beginning of the simulation, which has very limited impact on the interactions of nearby grids and the service level. Adding the parking zones also have no significant changes regarding the excess time. The major benefit of the parking zones is the reduction of rebalance mileages since the bikes are clustered to the zones and the vehicle only need to stop at the parking zones. Compared to the purely dockless systems, adding the zones can save rebalance mileage by 4%, 10%, and 11% in Los Angeles, Philadelphia, and Chicago, respectively (Figure 5.12 (c)). The change of zone capacity does not impact the rebalance mileages, since the grids that require rebalance stay nearly the same. However, with smaller zone capacity, there could be more grids exceeding the capacity during the day (Figure 5.12 (d)). For example, about 57 grids exceeded the 30-bike parking capacity in Chicago. The number of grids exceeding the capacity quickly drops when the zone capacity changes from 30 to 60, and becomes

near zero when the capacity is larger than 60. Therefore, a parking zone capacity around 60 works for most of the grids in all three cities, and Chicago can consider setting a larger capacity (e.g., 100) for the eight grids with very high parking needs. Overall, Chicago benefits the most from adding geo-fencing zones in a dockless BSS.

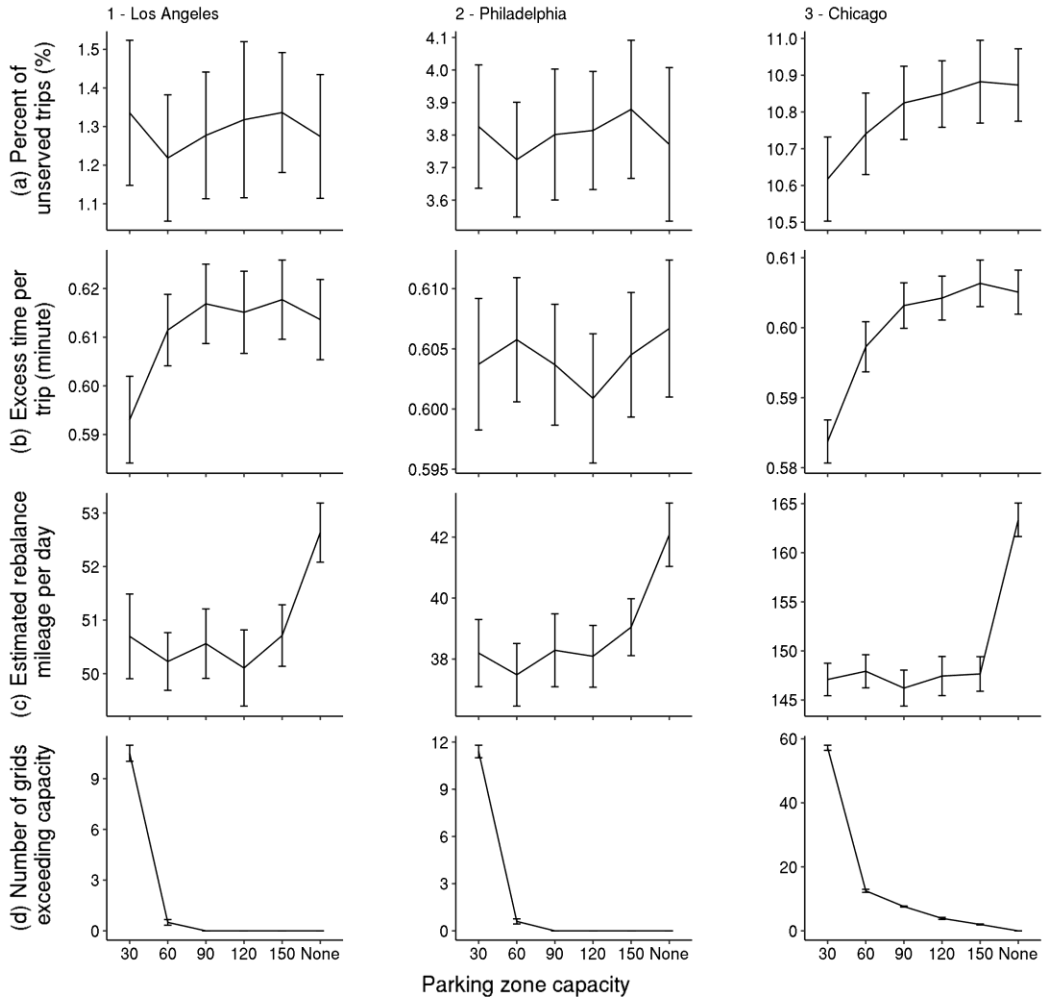


Figure 5.12. System performance when setting different capacity of the geo-fenced parking zones in a dockless system ("None" means there are no geo-fenced zones and the bikes are scattered)

5.4.3.2 The impact of bike supplies

This study alters the total number of bikes in the system by multiplying a ratio to the current bike supply and summarizes the change of the key metrics in Figure 5.13. In general, the conclusions regarding the performance of hybrid and dockless BSS stay consistent for all four metrics

(unserved trips, excess time, rebalance mileages, and bikes parked on the street). However, the performance of station-based BSSs is more sensitive to the bike supply, especially the percentage of unserved trips and the rebalance mileages. In Los Angeles and Chicago, when the bike supply ratio is smaller or equal to 0.8, the percentage of unserved trips in station-based BSSs becomes less than other types of BSSs (Figure 5.13 (a)), which indicates that station-based systems can serve more trips when the bike supply is insufficient. Nevertheless, reducing the bike supply also renders more trips becoming unserved, which is undesired. Adding more bikes can reduce the unserved trips but will come at the cost of increased excess time for station-based systems ((Figure 5.13 (b), worsened bike return issues) or increased on-street parking for dockless and hybrid systems ((Figure 5.13 (d)). Changing the bike supply also affects the rebalance (Figure 5.13 (c)). When the bike supply ratio is larger than 1.3 in Los Angeles and 1.5 in Philadelphia, a sudden drop of rebalance mileage can be observed. At these ratios, the bikes take up 64% and 67% of the total docking space, which limits the possible inventory variations of many stations, thus requiring less rebalance. But in Chicago, a bike supply ratio of 1.7 is needed for the bikes to take up 67% of the total docking space. That's why the reduced rebalance miles is not observed in Chicago.

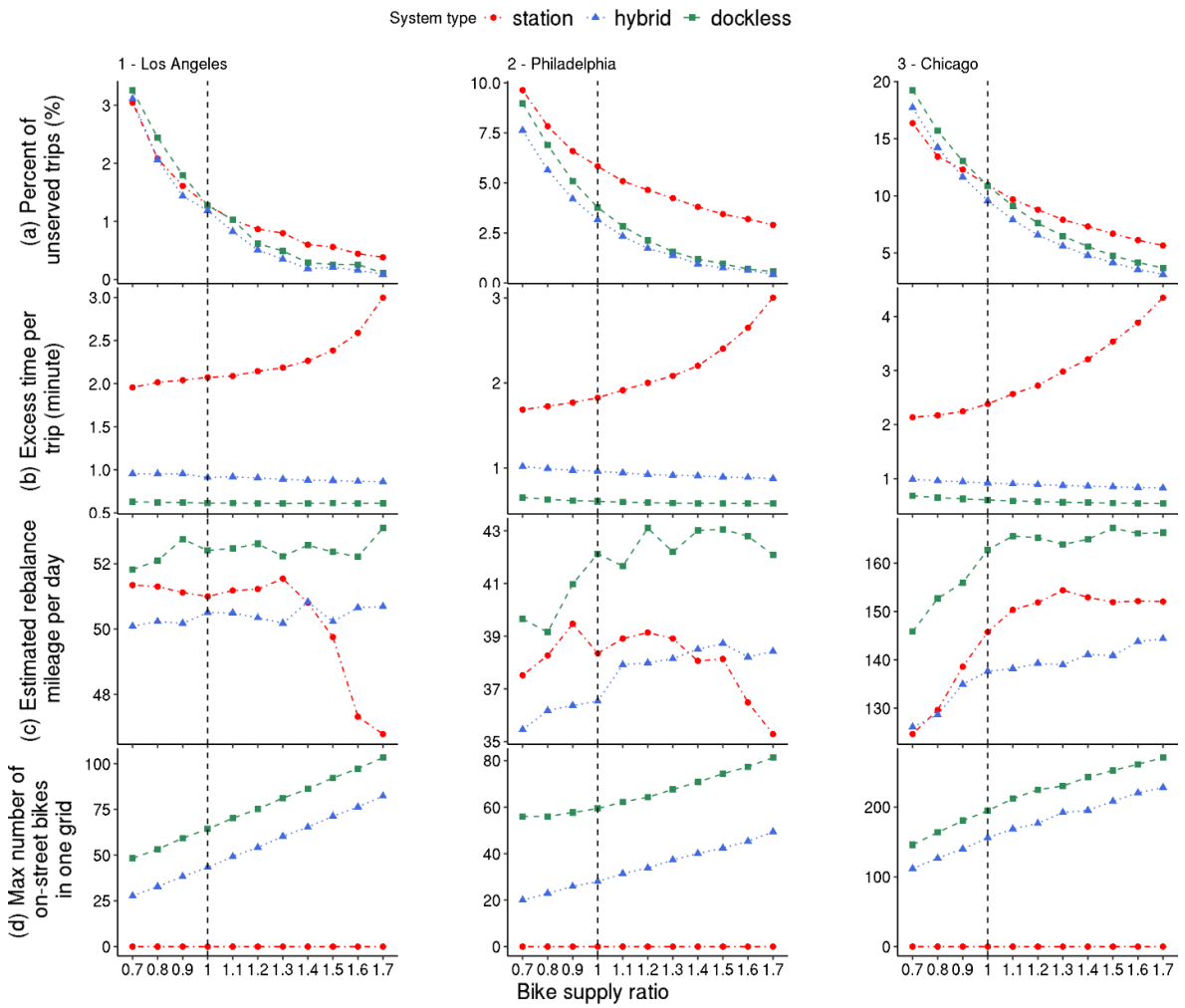


Figure 5.13. Results of sensitivity analysis of different bike supply (by multiplying the bike supply ratio to the current bike supply, which is shown by the vertical lines)

5.5 Conclusions and future work

In this study, a simulation modeling framework is proposed to evaluate different types of bike share systems from the aspects of service level, excess time, rebalance mileages, and on-street parking, which quantifies the user experience and the operational challenges in different types of BSSs. It is difficult to analyze the above aspects solely using bike share trip data. The user behaviors of accessing bikes may differ between users and will affect the interactions of neighboring grids. Therefore, system simulations are used to model such complexity. The proposed model first simulates the actual origins and destinations of bike share trips based on trip data from existing station-based systems. This study also proposes a user behavior model that

integrates historical travel survey data to model the walking behaviors of bike share users. In the case study, the model is applied to compare the above mentioned key metrics in station-based, dockless, and hybrid BSSs when serving the same demands. The BSSs in Los Angeles, Philadelphia, and Chicago are selected to evaluate the impact of system types to BSSs with different system sizes and usage intensities.

The simulation results show that, with the existing bike supply in the three analyzed cities, the hybrid BSS yields the best overall performance, because it reduces the unserved trips and rebalance mileages, and also has intermediate performance regarding the excess time and on-street parking bikes. One common trend observed from all three analyzed cities is that, dockless BSSs have the best user experience regarding excess time while station-based BSSs have the worst. The saving of time in a dockless BSS is because that users on average spend less time walking to a nearby bike when the bikes are scattered and also can directly park the bikes at the actual destinations. Hybrid systems have intermediate average excess time, which is closer to that of dockless BSSs. Overall, dockless BSSs could save 1.2 to 1.8 minutes excess time per served trip compared to the station-based BSSs, which corresponds to 10% to 15% of the median trip time in the three cities. The results of unserved trips are affected by the system's usage intensity and the spatial density of stations. In Los Angeles, where the usage intensity is low, the difference of unserved trips is not significant among different types of BSSs. However, in Philadelphia, both dockless and hybrid systems have less unserved trips than station-based BSSs. Compared to dockless BSS, the service level is further improved in hybrid systems in both Philadelphia and Chicago, due to having some users park the bikes to stations, where the demand is higher than surrounding grids. Different from Philadelphia, Chicago has similar unserved trips in the station-based and dockless BSSs. This is because the station-based BSS in Chicago has a denser station networks in high demand areas than Philadelphia, and the stations can better complement neighboring stations in the case of empty or full stations and redistribute bikes through forced rerouting (at the cost of user excess time). Due to the extra mileages needed to collect the scattered bikes, dockless BSSs require 4%, 16%, and 19% more rebalance mileages than the corresponding hybrid BSSs with a 50% park-to-station probability. The hybrid BSSs have similar rebalance mileages to station-based BSSs in both Los Angeles and Philadelphia, because the areas visited by the rebalance vehicles have similar spatial coverage. The rebalance vehicle traveled more mileages in the station-based BSS than in the hybrid

BSS in Chicago, due to the need to rebalance more outskirt grids. On-street parking could cause turmoil and safety issues if all bikes are randomly parked on the sidewalk. For the worst cases, one grid could have 64, 59, and 195 on-street bikes in Los Angeles, Philadelphia, and Chicago, respectively. The results show that, for the worst case with 195 dockless bikes in one grid, there are sufficient space to add geo-fencing zones to contain these bikes.

A sensitivity analysis on the geo-fencing zone capacity reveals that, the dockless BSSs benefited the most from adding geo-fencing zones in Chicago than in the other two cities. Besides addressing on-street parking issues, the geo-fencing zones also help reduce the rebalance workload, with a reduction of rebalance mileage up to 11% compared with the fully dockless BSSs. A parking zone capacity around 60 bikes is appropriate for most of the grids in the three cities. The sensitivity analysis of having different bike fleets shows that the conclusions regarding the performance of hybrid and dockless BSS stay consistent with different bike supplies. But the performance of station-based BSSs are more sensitive to the bike supply, especially for the percentage of unserved trips and the rebalance mileages.

Putting everything together, cities with relatively large station-based BSSs such as Philadelphia and Chicago may consider operating the BSS in a hybrid way and add geo-fencing zones in each grid to regulate parking behaviors. Philadelphia would benefit more from such transition regarding the service level since it has a relatively high demand and sparse station networks. For cities already have a station-based system but with a low usage intensity (e.g., Los Angeles), the benefit of transferring the existing station-based BSS to a dockless or hybrid system is limited.

While this study provides unique insights in comparing different types of BSSs, future research may improve the model from the following three aspects. First, the actual O-D estimation and user behavior model rely on the walking trip data from historical travel surveys, which assumes that bike share users' behaviors of walking to find a bike follow the same pattern as people walking to switch to another transportation mode in previous travel surveys. Obtaining such distance distribution from bike share users can help better model the user behaviors. Second, this study compared the system performance of different types of BSSs when serving the same demands. It is possible for the demand to change when the BSS type changes. Real-world data from before and

after system type transitions would be very precious to study such demand change. Third, taking into account the complex interactions of neighboring grids can improve the rebalance algorithms.

6. CONCLUSIONS

This dissertation develops various modeling frameworks and methods to improve the understanding of BSSs from the aspects of travel patterns, environmental benefits, system expansion, and the impacts of system types. The key findings and insights gained from the modeling results can improve policy making and BSS planning and operation.

The key contributions of the proposed models are summarized in Figure 6.1. This dissertation first analyzes the statistical property of the distance and duration of bike share trips (Chapter 2), which provides a fundamental basis for modeling BSSs use. In Chapter 3, a Bike Share Emission Reduction Estimation Model (BS-EREM) is proposed to quantify the environmental benefits of BSSs. The BS-EREM estimates bike share trips' transportation mode replacement with the consideration of heterogeneous travel mode choices. Trip distance, trip purpose, trip start time, and public transit accessibility are taken into account in the model. The BS-EREM can be applied to different cities to evaluate the environmental benefits of their BSSs, and thus support the cities' decisions for BSS development. In order to better take advantage of the benefits of the BSSs, this dissertation analyzes two potential ways to improve the BSSs: expanding the BSSs and changing the system types. For BSS expansion (Chapter 4), the piecewise regression model analyzes the station interactions in the BSS expansion process, which provides insights for the design of bike share station networks. In addition, a Spatial Eccentricity Quantile based Ensemble Model (SEQEM) is proposed, which improves the prediction performance of BSS expansion demand estimation by considering the spatial station network structure. The SEQEM also has the benefit of not depending on external socio-demographic and point-of-interest data. Therefore, the SEQEM can be easily applied to other cities only using historical bike share trip data and station location information. The improved prediction performance by SEQEM can help inform station sizing and bike allocation in the BSSs. To study the impact of changing system types (Chapter 5), a comprehensive stochastic simulation framework is proposed to evaluate the user experience and system operations in different types of BSSs. The simulation framework integrates a method to estimate the actual origins and destinations of bike share travel demands, a user behavior model, and a rebalance optimization model.

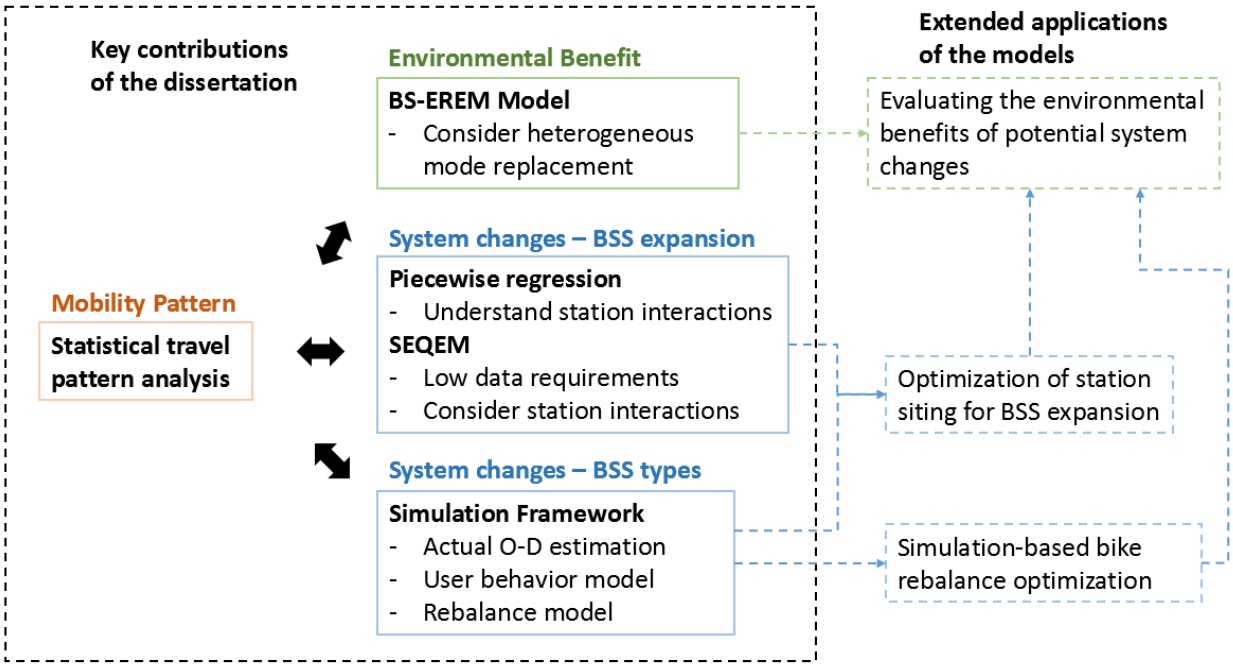


Figure 6.1. Key contributions of the dissertation and potential extended applications of the models

The models developed in this dissertation can be further integrated and extended in future research to better guide the development of BSSs (Figure 6.1). The SEQEM and the BSS simulation framework can be integrated to optimize new station siting for BSS expansion. Given a set of station location candidates, the bike share demands of a certain expansion plan can be estimated using the SEQEM. Then the estimated demand can be inputted into the BSS simulation framework to evaluate the system performances after the expansion. Finally, system designers can use the integrated model to identify the expansion plan that yields the best performances. Additionally, the BSS simulation framework can be modified into a simulation-based bike rebalance optimization model. In Section 5.4.2.1, the simulation results show that the rerouting behaviors of users can change the bike distributions and, thus, the fulfillment of demands. The bike allocation optimization model based on the expected demands did not consider such behaviors. In contrast, a simulation-based optimization model can capture such details and potentially improve the rebalance performance. Furthermore, for the optimized BSS expansion plan or the optimized bike rebalance from the abovementioned integrated models, the simulated trips can be inputted into the BS-EREM model to quantify the environmental benefits resulted from such optimized system changes. Overall, although the models and methods are developed specifically for bike share

systems, they are also transferable to other shared micro-mobility systems such as e-scooter share systems.

In the following sections of this chapter, the key findings and insights for BSS development are summarized in Section 6.1; then, the limitations and potential future research are discussed in Section 6.2.

6.1 Key findings and insights

The case studies using the proposed models in this dissertation generate the following key findings and insights:

(1) Smaller BSSs may expand the overall spatial coverage of the systems to better take advantage of BSSs' benefits. The spatial service coverage not only impacts the travel patterns, but also influences the environmental benefits of BSSs.

The travel pattern analysis in Chapter 2 reveals that, in larger systems (i.e., the BSSs in Washington DC, Chicago, and New York), both the trip distance and the trip duration follow a lognormal distribution. However, no common distributions have been found that can best describe the travel patterns of those smaller (regarding diameter and the number of stations) BSSs. The insufficient spatial coverage of the station network can restrict bike share users' movements to make long trips and, thus, affect the observed travel patterns. As a reference, the diameters (the longest distance between any two stations) of the BSSs in Washington DC, Chicago, and New York were 14.3, 23.3, and 11.2 miles, respectively, while the smaller system in Bay Area only had a diameter of 2.3 miles in 2016.

The overall spatial coverage of the stations also influences the per-trip environmental contributions of bike share trips. The larger systems have higher per trip emission reductions since the larger spatial coverage of the station networks allow users to make more long-distance trips, which are more likely to replace car trips. For example, the emission reduction per trip is 581 g CO2-eq in Chicago but only 283 g CO2-eq in the smaller BSS in Los Angeles. Such variations of unit emission reductions across different cities also indicate that it is risky to apply the per-trip emission

reduction learned in one city to another city for the environmental benefit estimation. The proposed BS-EREM model can account for such differences across different cities to more accurately estimate the environmental benefits of BSSs.

Therefore, the smaller BSSs should consider expanding the overall spatial coverage of service to enable more long trips. For example, in Chicago which had the largest system diameter in 2016 among the eight studied cities, the 95% percentile of trip distances is 3.2 miles. This can be used as a reference for expanding the smaller systems. One potential strategy is to set the distance from the high-demand areas (e.g., the downtown areas) to the system outer boundary greater than a distance threshold (e.g., 3.2 miles), so more users can travel between the high-demand areas and the outskirt areas.

(2) In station-based BSSs, the station density is an important factor that influences the ridership. The designing of station density should consider users' walking behaviors and station interactions.

The result of piecewise regression analysis on Chicago data in Chapter 4 indicates that existing stations receive maximum complement benefits when the nearest new stations are 0.3 miles away. In this case, local residents and visitors find that there are more options on where to pick up and return bikes, which can potentially attract more new users and encourage the usage of existing users. When the new station is less than 0.11 miles away from an existing station, the competition between them is more prevalent. The 0.11 miles correspond to the 89% percentile of the walking distances in the CHTS survey when people use walking to transit to other modes (Section 5.3.2.2). Therefore, the new stations dilute the demands of existing stations because the users find that there are more options within the walkable distance range. In such situations, adding the new stations may not lead to significant increases of demands. The system designers need to more carefully evaluate the trade-off between the station installation cost and the potential benefits to the ridership of whole BSSs.

It is possible that the above discussed distance thresholds (0.3 miles and 0.11 miles) could vary across different cities. The piecewise regression model can be applied to other cities to identify the city-specific patterns of station interactions in different cities. The proposed SEQEM model can

also improve the demand estimation for BSS expansion planning by considering the spatial structure of station networks. Both the piecewise regression model and SEQEM only require bike share trip count data and station location information (latitude and longitude), which can be easily transferred to other cities.

(3) Although dockless and hybrid BSSs have emerged in many cities in recent years, they may not yield the best performances for all the cities. The important factors impacting system performances include demand intensity, station density (for station-based and hybrid BSSs), and user behaviors.

In Philadelphia and Chicago, where the overall demand is higher than Los Angeles, the hybrid system has the best overall performance because it has the lowest unserved trips and rebalance mileage and has the intermediate performance in terms of excess time and on-street-parking issues. In contrast, the difference in overall performances among different system types is less significant in cities with low usage intensity such as Los Angeles.

The station density also affects the differences of service levels between a station-based BSS and the corresponding hybrid BSS. In Section 5.4.2.1, the analysis shows that, compared with Chicago, Philadelphia can benefit more regarding the service level from transitioning the BSS type from station-based into hybrid. This is because Chicago has a relatively denser existing station network in the high-demand areas, which allows the users to more easily reroute to nearby stations when the first-choice stations are empty or full.

In addition, the user rerouting behaviors can indirectly affect system performances. Although station capacities restrict bike allocations in station-based BSSs, forced rerouting trips (returning bikes to neighboring stations when target stations are full) indirectly rebalance systems. This benefit is lost in dockless systems, partially canceling the benefit of bike allocation flexibility (Section 5.4.2.3).

Overall, the selection of BSS types should take into account many complex factors such as demand intensity, station density (for station-based and hybrid BSSs), and the user behaviors. The proposed BSS simulation framework in Chapter 5 is capable to better support the decisions by considering

these factors. Similar to the three case-study cities in Chapter 5, other cities that already have a station-based BSS can just follow the same procedures outlined in Section 5.3 to apply the simulation model. In Section 5.3.3.4, this dissertation has discussed how to modify the demand estimation process in order to run the model using dockless or hybrid system data as the input.

6.2 Limitations and Future Research Directions

The limitations of this dissertation and potential future research directions are outlined below:

First, more data about the travel behaviors of bike share users is needed to improve the modeling of BSSs. For the environmental benefit estimation, the case study results are based on the assumption that the BSS users have the same travel pattern as an average person in the urban area of each city. Similarly, the user behavior model in the BSS simulation framework also assumes that the walking distance preference of BSS users follow the same distribution in historical travel surveys. To address such issues, more data should be collected specifically from BSS users. For example, to improve the BS-EREM, mode substitution information for a specific bike share trip should be collected using user surveys, while more detailed trip information such as trip distance, purpose and start time should also be recorded. Researchers may also hire volunteers from BSS users and use wearable GPS devices to track their walking behaviors when accessing a bike. Such information will significantly improve the user behavior model.

Second, the differences in ridership, environmental benefits, and system performances among the BSSs in different cities are also affected by some external factors, which require more analysis in future research. Such factors include travel cost, weather, terrain, safety, and so on.

- Travel cost is one of the important factors that influence people's transportation mode choices (Buehler & Hamre, 2015). Different pricing schemes in different BSSs could impact the popularity of bike share, which will then influence the system performance and the selection of best-performance system types. Analyzing the travel cost of bike share should also consider whether bike share has competitive travel cost in the context of all transportation modes. Taking into account the heterogeneous travel cost preference of bike share users will improve the modeling of mode substitution in the BS-EREM. However, this would require the collection of detailed travel cost information in future travel surveys.

In addition, when studying the demand changes in the expansion process, future studies should also consider the changes of pricing schemes, which can potentially lead to irregular demand change patterns.

- Both the weather and terrain in different cities can impact the overall bike share ridership and the user behaviors. The cold winter weather or hot summer weather in certain cities could significantly reduce the usage of bike share. In some cities with undulating terrains, it would be more difficult to ride a bike to go up a steep street. In this case, the e-bikes could be a more popular option. In addition, both the weather and terrain conditions can also affect the users' walking preference when picking up a bike. Intuitively, users' acceptable walking distance will be reduced in the case of extreme weather and steep streets. Therefore, future studies should also consider the weather and terrain conditions when collecting user travel behavior data and modeling walking behaviors. Moreover, the designers of station-based BSSs may adjust the station density according to the terrains and the common weather conditions in different cities to ensure the stations are within a walkable distance for bike pick-up.
- Safety is another concern when riding a bike. The availability of bike lanes and other biking-friendly infrastructures can influence the usage of bike share (Buck & Buehler, 2012). In addition, a higher crime rate in certain areas is often related to lower mode shares of walking and biking (Joh et al., 2012), which should also be considered to improve the user behavior modeling.

Third, the demand prediction models for large-scale expansion need to be improved. The prediction results for BSS expansion show that the models perform poorly for the large-scale expansion in 2015 in Chicago. With a 58% increase of the number of stations, the patterns learned from historical data are insufficient to make desirable predictions. One potential method to improve this is to learn the experience from the BSSs in other cities that had large-scale expansion. So far, the prediction models that transfer the knowledge learned from one city to another city are very limited (Liu et al., 2018) and should receive more attention in future research.

Fourth, different BSS types' impact on the bike share demands should be further studied. Chapter 5 applies the proposed simulation framework to compare different types of BSSs when serving the

same demands; however, the system types could also impact the demands. One option to improve this is to assume different scenarios of demand changes, and apply the proposed simulation model with various demand scenarios to learn about the differences of system performances. Another alternative is to conduct surveys, asking people whether they will increase and decrease the usage of bike share and by how much if the system changes to a certain type. It is also important to collect before-and-after data from cities that changed the BSS types. So far, such cases are still very rare.

Lastly, as pointed out in Figure 6.1, the proposed models in this dissertation can be extended and integrated to further address additional problems for BSS development. For instance, the SEQEM can be combined with the BSS simulation framework to evaluate the potential system performances for different BSS expansion plans; the BSS simulation framework can be extended to be a simulation-based bike rebalance optimization model; the simulated trips from these extended models can then be inputted into the BS-EREM model to study the potential environmental benefits from such optimized system changes.

In summary, BSSs have great potential to reduce carbon emissions and can be further improved to play a more important role in sustainable transportation systems. The spatial service coverage of BSSs is very important for taking advantage of the benefits of BSSs. For designing the station networks of station-based BSSs, the planners should at least guarantee that the overall coverage of the station networks does not restrict the users' travel using bike share. This allows users to make more long-distance trips, which will replace more car trips and, thus, lead to more GHG emission reductions. The improved spatial service coverage also would complement the demands of existing stations. The proposed SEQEM and BSS simulation framework are very useful tools to assist the BSS expansion planning, while the BS-EREM helps estimate the environmental benefits of BSSs. The BSS simulation framework can provide customized suggestions for the system type selection for a city. The simulation results also exhibit the complex interactions between the users, stations and bikes, which indicate the necessity to apply simulation-based models for BSS modeling. Studying the user behaviors is very important for both long-term planning and short-term operations in BSSs. Future studies should collect more data, specifically from bike share users, and develop more realistic user behavior models to guide the development and operations of BSSs.

The shared micro-mobility options such as bike share and e-scooter share are very promising transportation modes and can contribute more to the sustainability of urban transportation systems. Although the total ridership of bike share in the United States has kept increasing from 2010 to 2019, bike share is also facing competition from e-scooter share. E-scooter share has only gained popularity in the past two years. However, the ridership of e-scooter share had increased dramatically from 2018 to 2019, and has just exceeded the ridership of bike share in 2019 (NACTO, 2020). Therefore, future studies should also pay more attention to e-scooter share. The models proposed in this dissertation can also be modified to model e-scooter systems.

APPENDIX A. SUPPLEMENTAL INFORMATION FOR CHAPTER 2

A.1 Trip data

Table A. 1. A sample of the trip data from Seattle's Pronto Bike Share program

Trip id	Start time	Stop time	Bike id	Trip duration (seconds)	Start station name	End station name	Start station id	End station id
431	10/13/2014 10:31	10/13/2014 10:48	SEA00298	985.935	2nd Ave & Spring St	Occidental Park / Occidental Ave S & S Washington St	CBD-06	PS-04
432	10/13/2014 10:32	10/13/2014 10:48	SEA00195	926.375	2nd Ave & Spring St	Occidental Park / Occidental Ave S & S Washington St	CBD-06	PS-04
518	10/13/2014 12:20	10/13/2014 12:31	SEA00321	690.793	City Hall / 4th Ave & James St	2nd Ave & Blanchard St	CBD-07	BT-05
519	10/13/2014 12:20	10/13/2014 12:26	SEA00089	351.563	1st Ave & Marion St	2nd Ave & Spring St	CBD-05	CBD-06
520	10/13/2014 12:21	10/13/2014 12:32	SEA00028	653.071	Burke Museum / E Stevens Way NE & Memorial Way NE	UW Intramural Activities Building	UW-02	UW-07
528	10/13/2014 12:39	10/13/2014 12:45	SEA00415	357.776	6th Ave S & S King St	Occidental Park / Occidental Ave S & S Washington St	ID-04	PS-04
530	10/13/2014 12:43	10/13/2014 12:48	SEA00311	278.849	King Street Station Plaza / 2nd Ave Extension S & S Jackson St	King Street Station Plaza / 2nd Ave Extension S & S Jackson St	PS-05	PS-05

A.2 Data Processing for Bay Area's Ford Gobike program

Bay Area bike share system (Figure A. 1) includes multiple sub-systems: San Francisco, San Jose and several other cities or towns. In this case, only the intra-city trips in San Francisco, which makes up 89.7% of the total trips in Bay Area, are analyzed. There exists only four inter-city trips in Bay Area in 2016.

Table A.2 shows the 4 inter-city trips in Bay Area bike share system. Three of the four trips have trip duration that is only 1/3 of the duration estimated by Google Maps API. These trips are much shorter in time than normal bike trips. It is highly likely that the bike was carried by a vehicle for most part of the trips. In addition, such trips are relatively trivial compared to the total trips (983,648 trips) in Bay Area. Therefore, these trips are treated as outlier and removed for the analysis.

Figure A. 1. Bike share stations in Bay Area

Table A.2. Inter-city trip record of Bay Area bike share system in 2016

duration (min)	start_time	end_time	start_station_id	start_station_lat	start_station_lon	end_station_id	end_station_lat	end_station_lon	bikeid	usertype	estimated distance (mile)[1]	estimated duration (min)[2]
94	6/23/2016 18:46	6/23/2016 20:20	65	37.77106	-122.403	2	37.32973	-121.902	43	Subscriber	55.43266	292.0333
488.05	7/25/2016 8:41	7/25/2016 16:49	56	37.79225	-122.397	3	37.3307	-121.889	295	Subscriber	57.5329	305.3167
88.78333	3/8/2016 21:38	3/8/2016 23:06	27	37.38922	-122.082	57	37.78175	-122.405	132	Subscriber	39.41927	210.0833
94.7	2/4/2016 18:12	2/4/2016 19:47	11	37.33589	-121.886	70	37.77662	-122.395	696	Subscriber	55.6054	290.2333

Notes:

1. Trip distance estimated by Google Maps API
2. Trip duration estimated by Google Maps API

A.3 Model Selection (all trips)

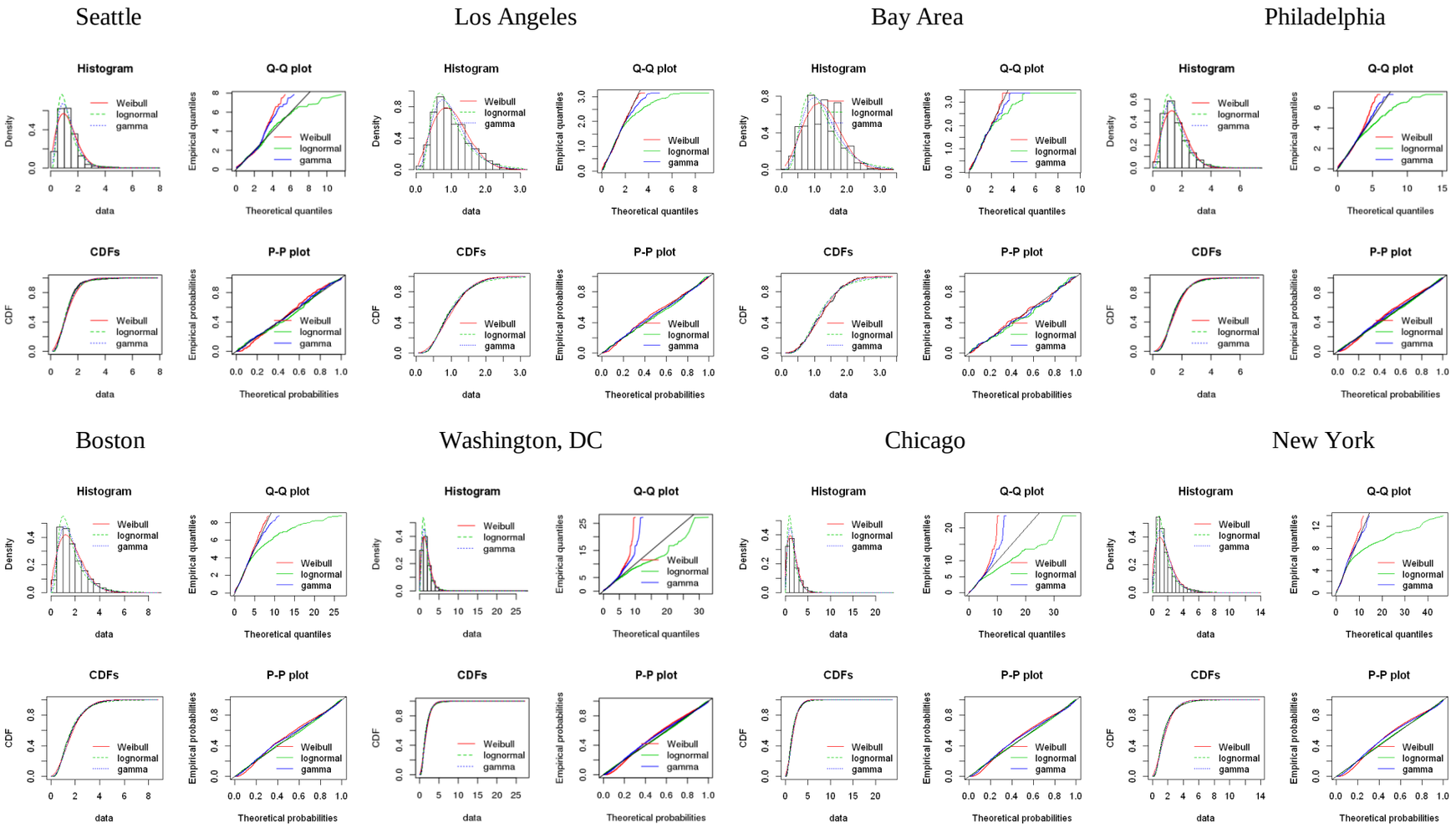


Figure A. 2. Goodness of fit plots for the 8 cities (commuting trip distance)

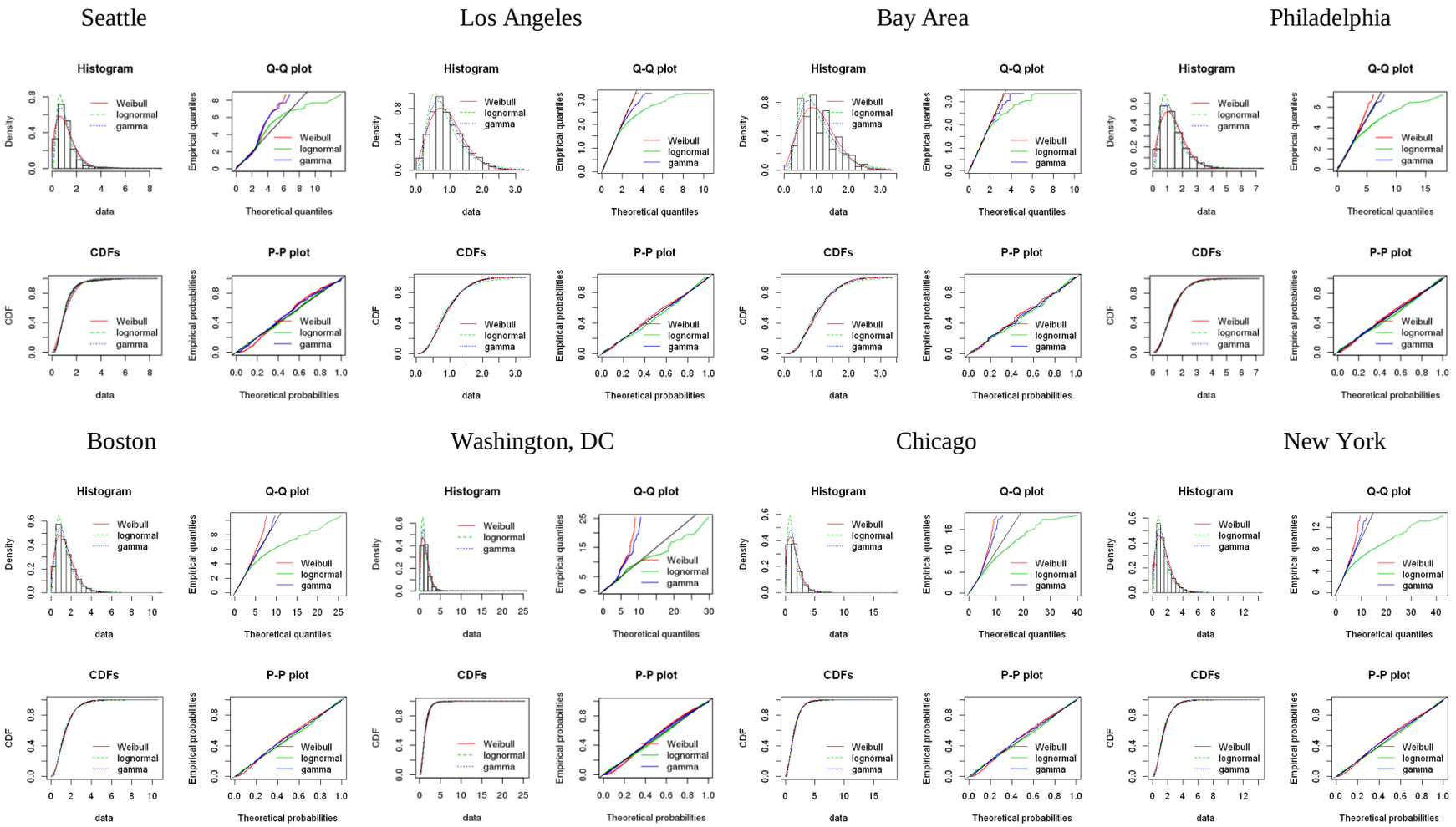


Figure A. 3 . Goodness of fit plots for the 8 cities (tourist trip distance)

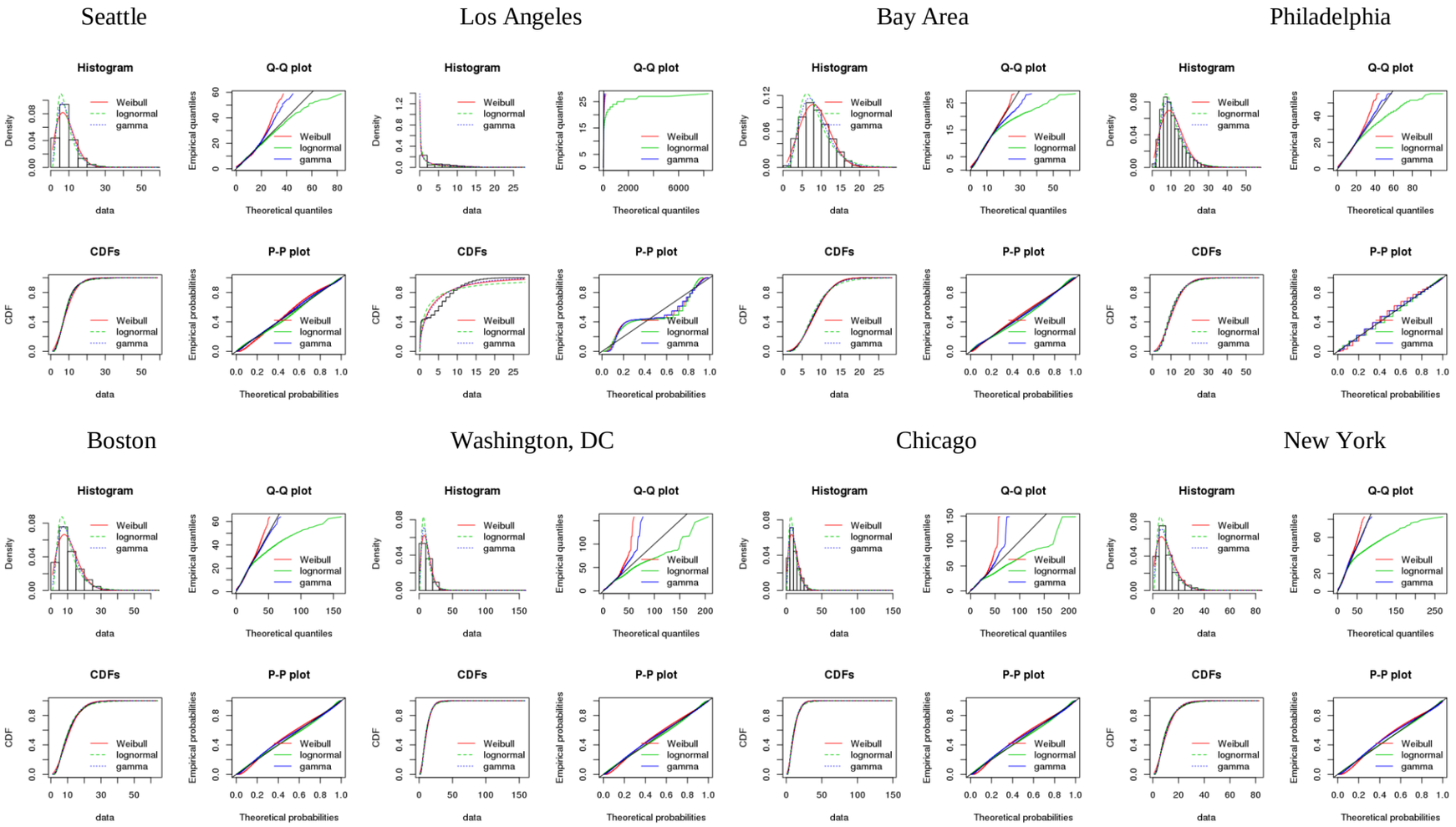


Figure A. 4. Goodness of fit plots for the 8 cities (commuting trip duration)

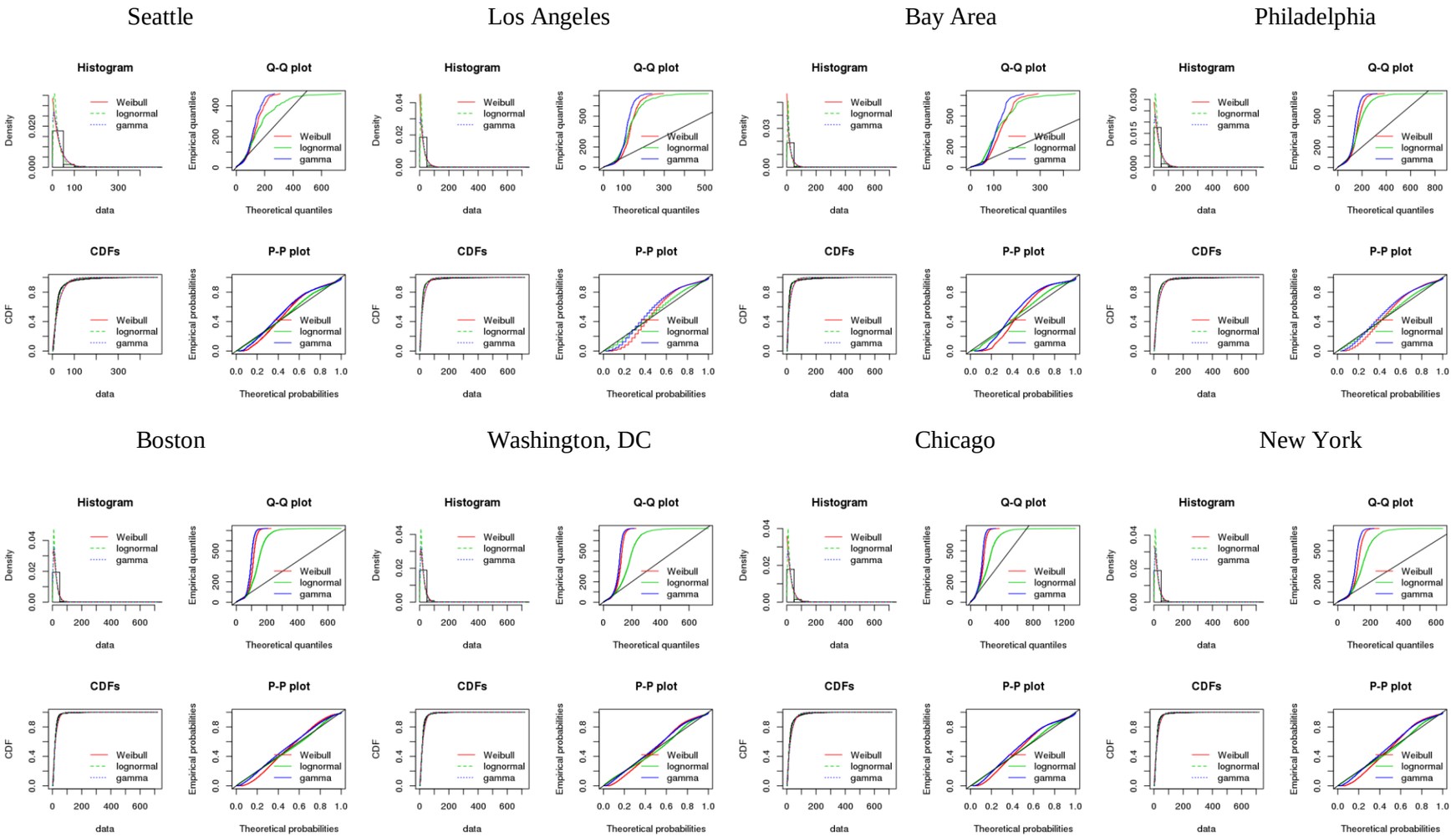


Figure A. 5. Goodness of fit plots for the 8 cities (tourist trip duration)

Table A. 3. Results of model selection for touristic trip distance

			Seattle	**Los Angeles**	**Bay Area**	**Philadelphia**	**Boston**	**Washington, DC**	**Chicago**	**New York**
Weibull	Parameters	Shape	1.5341	1.9885	2.0785	1.8425	1.6653	1.5253	1.5056	1.5661
		Scale	1.2874	1.0580	1.1936	1.5458	1.6037	1.5650	1.7359	1.6558
	Goodness-of-fit statistics/criteria	K-S statistic	0.0816	**0.0349***	0.0555	**0.0356***	**0.0465***	0.0507	**0.0443***	**0.0436***
		AIC	77748.5	78612.7	67970.5	311526.0	873503.2	2447251.5	3585888.8	9457730.1
		BIC	77765.7	78630.7	67988.0	311545.8	873524.9	2447275.2	3585913.1	9457756.4
Gamma	Parameters	Shape	2.6119	3.3305	3.6507	3.0787	2.6219	2.4627	2.2857	2.3599
		Scale	2.2745	3.5600	3.4622	2.2516	1.8400	1.7599	1.4700	1.5961
	Goodness-of-fit statistics/criteria	K-S statistic	0.0621	**0.0186***	**0.0350***	**0.0190***	**0.0312***	**0.0304***	**0.0365***	**0.0297***
		AIC	73259.6	**78176.4**	**67349.9**	**305807.4**	**855427.0**	2358109.0	3499519.1	9274230.0
		BIC	73276.7	**78194.4**	**67367.4**	**305827.2**	**855448.7**	2358132.7	3499543.3	9274256.2
Lognormal	Parameters	meanlog	-0.0651	-0.2242	-0.0901	0.1418	0.1515	0.1194	0.2070	0.1645
		sdlog	0.6301	0.5991	0.5670	0.6099	0.6605	0.6686	0.7005	0.6978
	Goodness-of-fit statistics/criteria	K-S statistic	**0.0335***	0.0505	0.0695	0.0364	0.0312	**0.0291***	**0.0257***	**0.0205***
		AIC	**69738.9**	82975.4	70573.5	309903.3	858637.9	**2319503.7**	**3459210.7**	**9271269.9**
		BIC	**69756.1**	82993.5	70591.0	309923.1	858659.5	**2319527.3**	**3459235.0**	**9271296.2**
Best Fit Distribution			Lognormal	Gamma	Gamma	Gamma	Gamma	Lognormal	Lognormal	Lognormal
Parameters			$\mu = -0.07$	$\alpha = 3.33$	$\alpha = 3.65$	$\alpha = 3.08$	$\alpha = 2.62$	$\mu = 0.12$	$\mu = 0.21$	$\mu = 0.16$
			$\sigma = 0.63$	$\beta = 3.56$	$\beta = 3.46$	$\beta = 2.25$	$\beta = 1.84$	$\sigma = 0.67$	$\sigma = 0.70$	$\sigma = 0.70$

1. * Significant at 0.05.
2. Lowest AIC and BIC scores are bolded when the K-S statistics are significant at 0.05.
3. Model parameters corresponds to the parameters in Table 3.

Table A. 4. Results of model selection for commuting trip duration

			Seattle	**Los Angeles**	**Bay Area**	**Philadelphia**	**Boston**	**Washington, DC**	**Chicago**	**New York**
Weibull	Parameters	Shape	1.8550	0.6329	2.5156	2.0402	1.8383	1.7164	1.7394	1.6145
		Scale	10.0013	3.4294	9.6507	12.4873	12.2131	12.5208	12.3974	12.2049
	Goodness-of-fit statistics/criteria	K-S statistic	0.0586	0.1792	**0.0227***	0.0812	**0.0471***	0.0402	**0.0439***	0.0582
		AIC	320810.9	486983.9	763833.0	1931294.6	5184440.4	9124747.4	13275482.9	40357869.5
		BIC	320828.8	487003.0	763852.7	1931315.9	5184463.6	9124771.7	13275508.0	40357896.8
Gamma	Parameters	Shape	3.4615	0.5112	5.1468	3.9289	3.1412	2.8342	2.8419	2.5688
		Scale	0.3917	0.1118	0.6021	0.3565	0.2911	0.2554	0.2588	0.2366
	Goodness-of-fit statistics/criteria	K-S statistic	**0.0296***	0.1770	**0.0200***	0.0606	**0.0355***	**0.0279***	**0.0328***	0.0449
		AIC	315744.1	482619.4	**763044.5**	1909056.8	5139295.2	9038005.0	13166770.2	39941250.2
		BIC	315762.0	482638.5	**763064.2**	1909078.1	5139318.4	9038029.4	13166795.3	39941277.5
Lognormal	Parameters	meanlog	2.0276	0.2821	2.0455	2.2671	2.2113	2.2202	2.2101	2.1776
		sdlog	0.5589	1.9774	0.4683	0.5274	0.5928	0.6267	0.6266	0.6530
	Goodness-of-fit statistics/criteria	K-S statistic	**0.0319***	0.2191	**0.0451***	0.0558	**0.0259***	**0.0249***	**0.0274***	**0.0133***
		AIC	**314757.4**	495707.0	771098.6	1910426.3	**5133890.4**	**9024751.9**	**13152295.7**	**39708472.4**
		BIC	**314775.2**	495726.1	771118.3	1910447.6	**5133913.6**	**9024776.3**	**13152320.7**	**39708499.7**
Best Fit Distribution			Lognormal	None	Gamma	None	Lognormal	Lognormal	Lognormal	Lognormal
Parameters			$\mu = 2.03$	-	$\alpha = 5.15$	-	$\mu = 2.21$	$\mu = 2.22$	$\mu = 2.21$	$\mu = 2.18$
			$\sigma = 0.56$	-	$\beta = 0.60$	-	$\sigma = 0.59$	$\sigma = 0.62$	$\sigma = 0.63$	$\sigma = 0.65$

1. * Significant at 0.05.
2. Lowest AIC and BIC scores are bolded when the K-S statistics are significant at 0.05.
3. Model parameters corresponds to the parameters in Table 3.

Table A. 5. Results of model selection for touristic trip duration

			Seattle	**Los Angeles**	**Bay Area**	**Philadelphia**	**Boston**	**Washington, DC**	**Chicago**	**New York**
Weibull	Parameters	Shape	1.0101	0.9630	0.9058	1.0148	1.0964	1.0127	1.1820	1.1689
		Scale	28.1208	23.0748	19.7892	31.8312	23.4076	26.1706	23.4318	21.4678
	Goodness-of-fit statistics/criteria	K-S statistic	0.1198	0.1560	0.1642	0.1241	0.1035	0.1051	0.0937	0.0922
		AIC	338614.0	503179.6	373796.1	1288168.9	3030850.9	8679039.4	11038732.2	30052685.3
		BIC	338631.2	503197.6	373813.6	1288188.6	3030872.6	8679063.0	11038756.5	30052711.6
Gamma	Parameters	Shape	1.2389	1.2175	1.0775	1.3135	1.5416	1.2372	1.7200	1.7735
		Scale	0.0443	0.0516	0.0508	**0.0416***	0.0688	**0.0476***	0.0786	0.0883
	Goodness-of-fit statistics/criteria	K-S statistic	0.1292	0.1593	0.1812	0.1302	0.0939	0.1108	0.0781	0.0585
		AIC	337571.5	502018.5	374872.7	1282093.2	3000916.1	8652275.5	10912615.5	29628874.6
		BIC	337588.7	502036.6	374890.2	1282112.9	3000937.7	8652299.2	10912639.8	29628900.8
Lognormal	Parameters	meanlog	2.8760	2.6980	2.5228	3.0258	2.7512	2.8026	2.7680	2.6927
		sdlog	0.8857	0.8246	0.8455	0.8310	0.7904	0.8988	0.7740	0.7458
	Goodness-of-fit statistics/criteria	K-S statistic	0.0527	0.0671	0.0736	**0.0496***	**0.0410***	**0.0406***	**0.0400***	**0.0291***
		AIC	326305.5	474645.8	349130.3	**1232427.7**	**2909092.5**	**8387075.5**	**10694832.5**	**28879473.4**
		BIC	326322.6	474663.8	349147.8	**1232447.5**	**2909114.1**	**8387099.2**	**10694856.8**	**28879499.7**
Best Fit Distribution			None	None	None	Lognormal	Lognormal	Lognormal	Lognormal	Lognormal
Parameters			-	-	-	$\mu = 3.03$	$\mu = 2.75$	$\mu = 2.80$	$\mu = 2.77$	$\mu = 2.69$
			-	-	-	$\sigma = 0.83$	$\sigma = 0.79$	$\sigma = 0.90$	$\sigma = 0.77$	$\sigma = 0.75$

1. * Significant at 0.05.
2. Lowest AIC and BIC scores are bolded when the K-S statistics are significant at 0.05.
3. Model parameters corresponds to the parameters in Table 3.

A.4 Power law and Exponential fitting for long trips

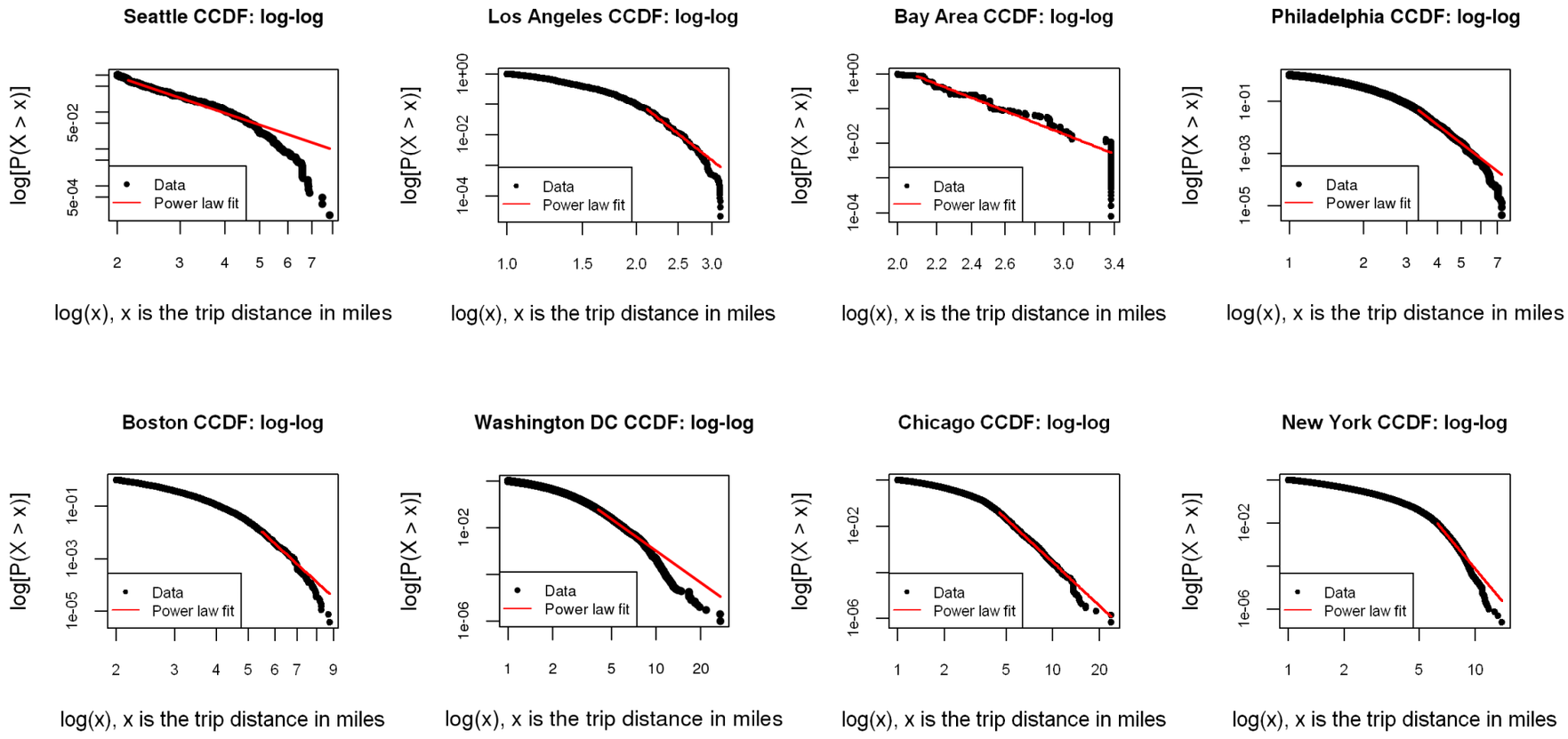


Figure A. 6. Power Law fit to the commuting trip distance

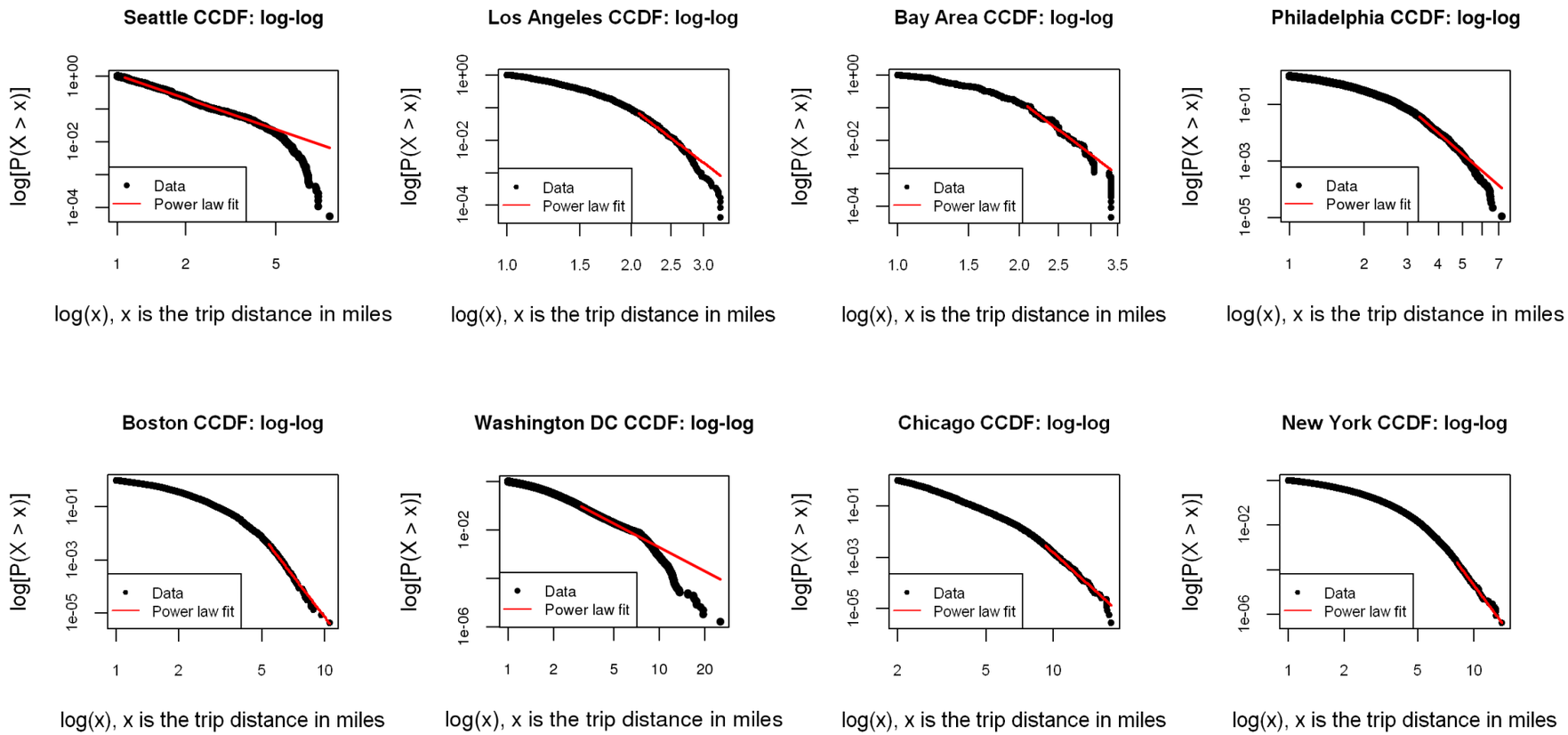


Figure A. 7. Power Law fit to the touristic trip distance

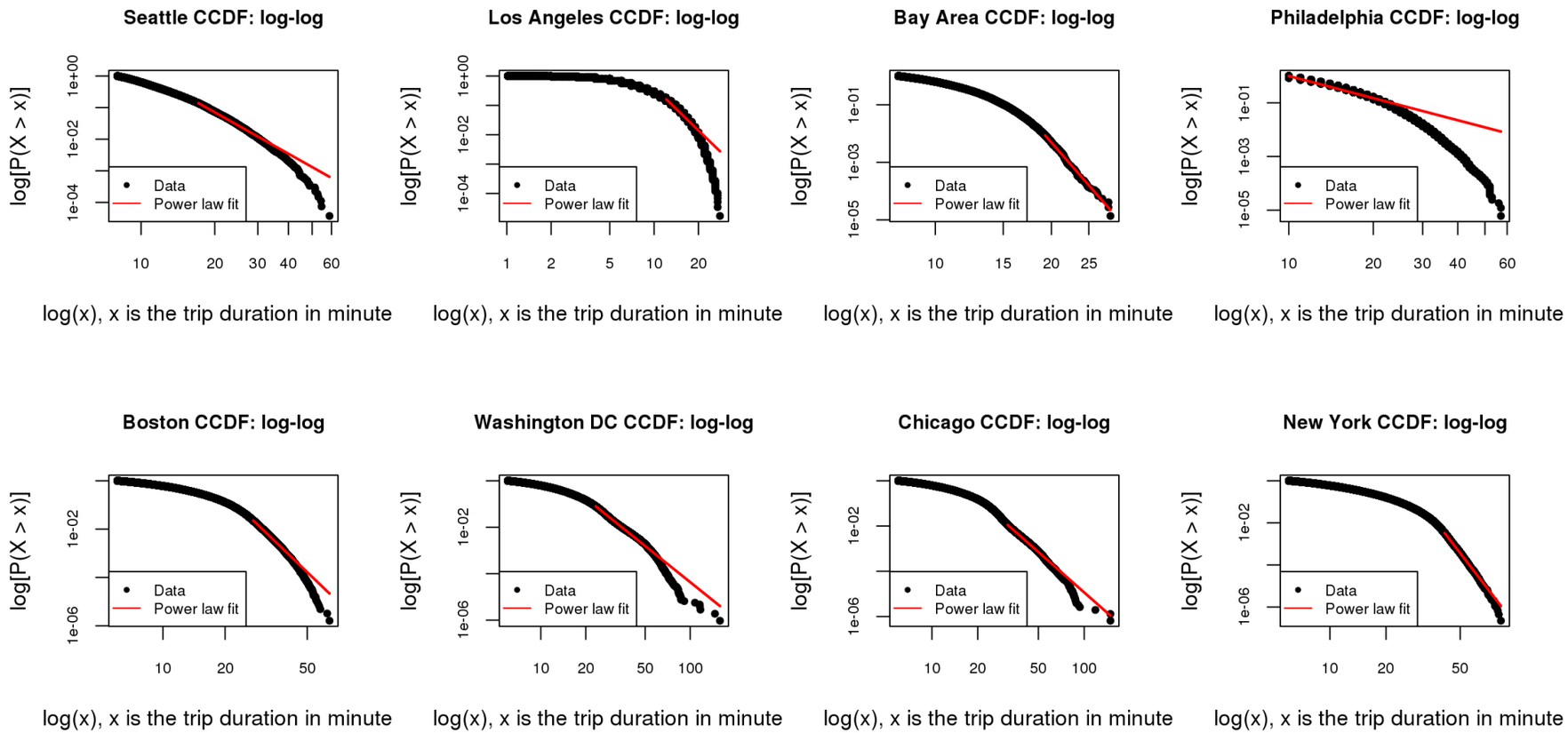


Figure A. 8. Power Law fit to the commuting trip duration

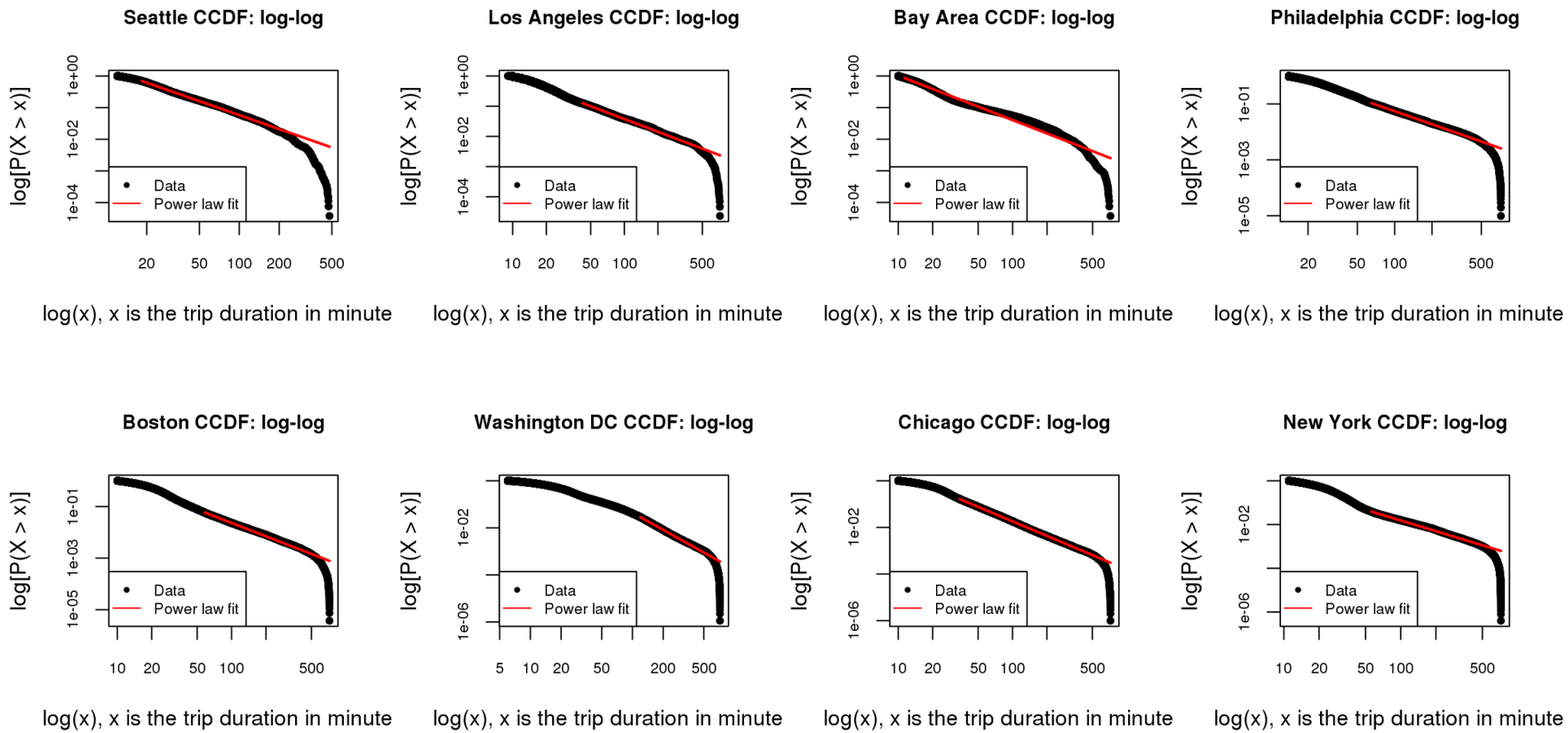


Figure A. 9. Power Law fit to the commuting trip duration

Table A. 6. Results of exponential distribution fitting

		Seattle	Los Angeles	Bay Area	Philadelphia	Boston	Washington, DC	Chicago	New York
Commuting trip distance (mile)	Rate parameter λ	0.3621	0.6764	0.4392	0.5386	0.3389	0.4657	0.4569	0.4551
	K-S statistic	0.5158	0.4920	0.5848	0.4166	0.4923	0.3724	0.3669	0.3657
Tourist trip distance(mile)	Rate parameter λ	0.5809	0.6884	0.6544	0.5545	0.5169	0.5234	0.3273	0.4935
	K-S statistic	0.4411	0.4978	0.4806	0.4258	0.4038	0.4076	0.4804	0.3897
Commuting trip duration (minute)	Rate parameter λ	0.0795	0.1254	0.0878	0.0663	0.0772	0.0745	0.0747	0.0741
	K-S statistic	0.4707	0.2965	0.5052	0.4846	0.3714	0.3611	0.3620	0.3598
Tourist trip duration (minute)	Rate parameter λ	0.0265	0.0328	0.0314	0.0243	0.0352	0.0342	0.0366	0.0374
	K-S statistic	0.2728	0.2795	0.2702	0.2883	0.2968	0.1862	0.3069	0.3380

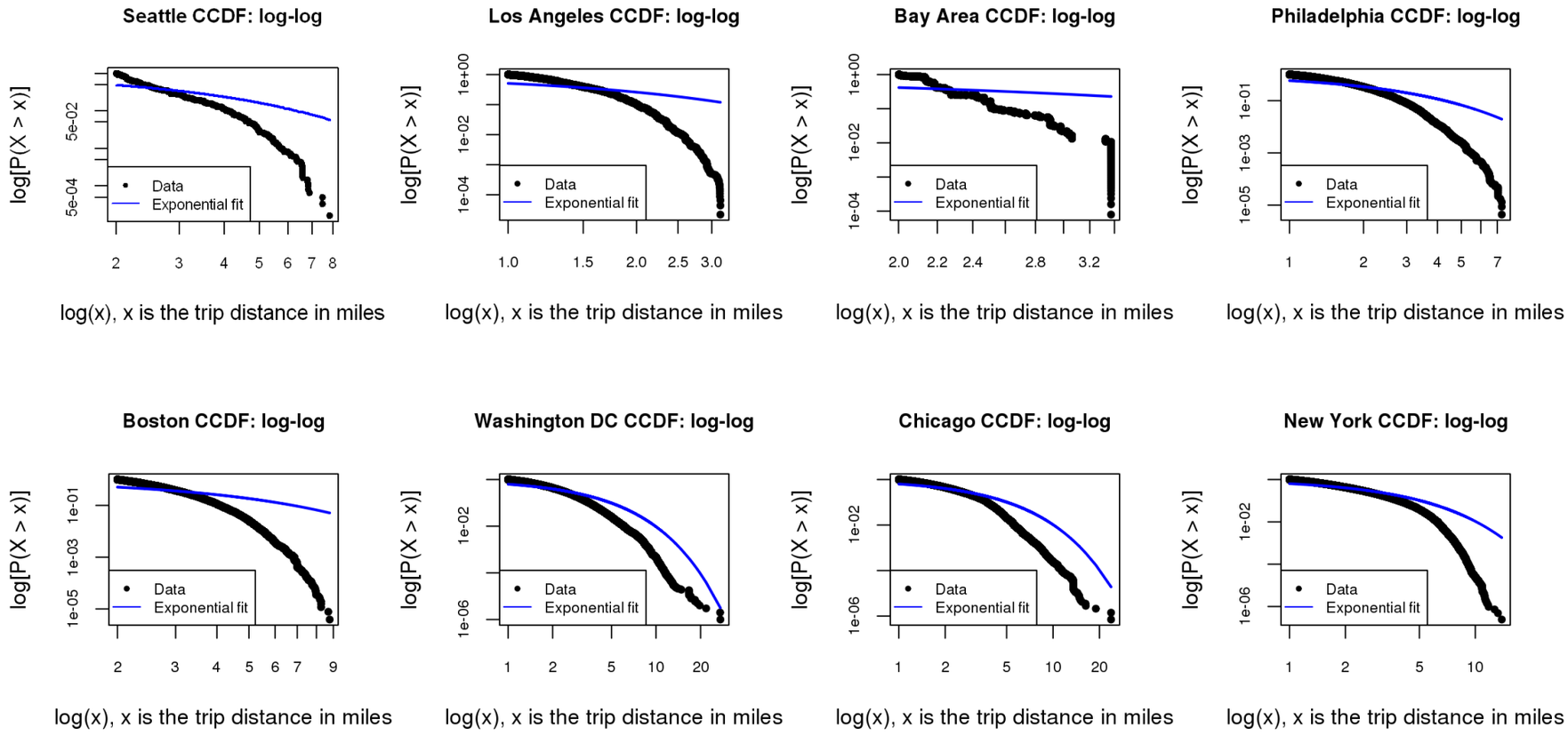


Figure A. 10. Exponential fit to the commuting trip distance

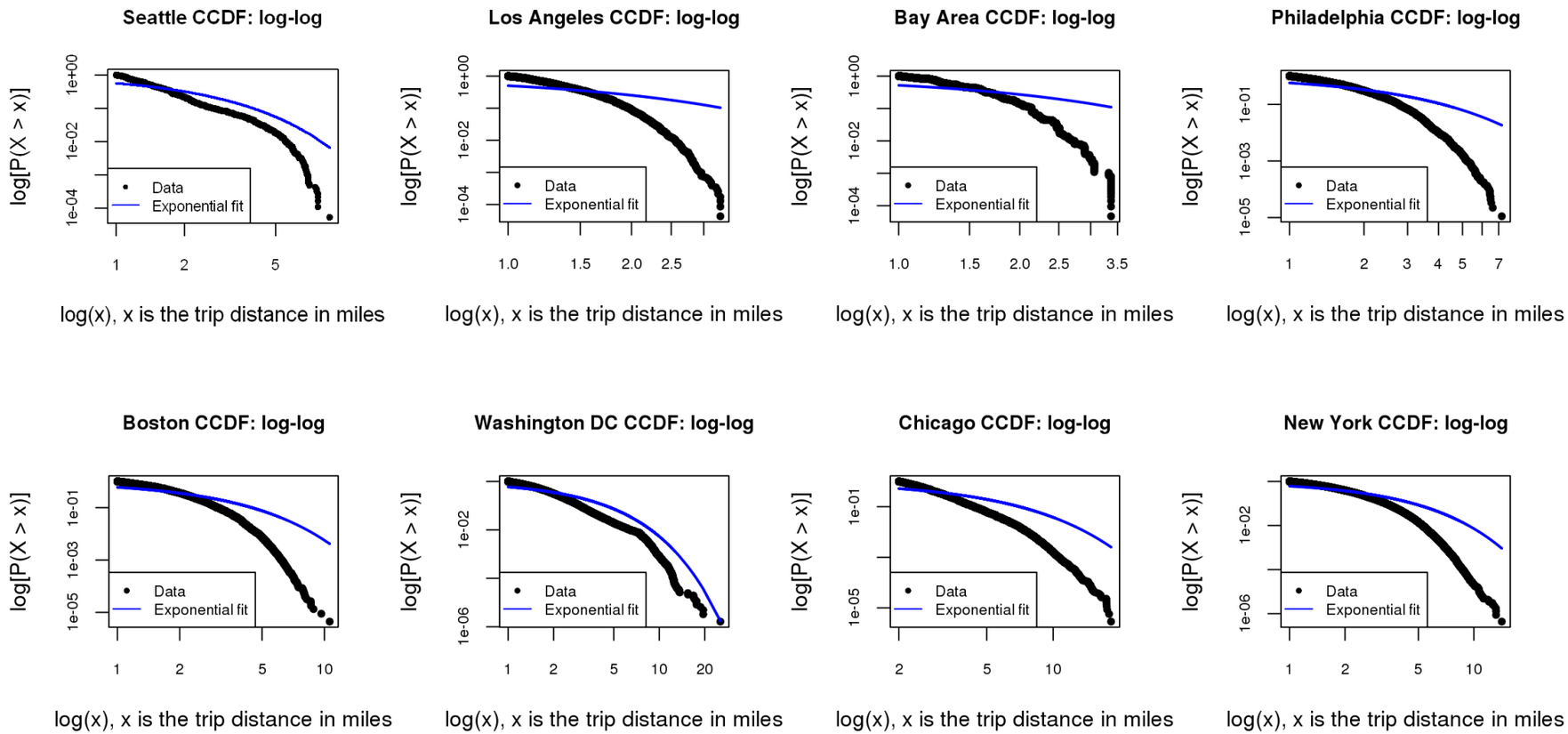


Figure A. 11. Exponential fit to the touristic trip distance

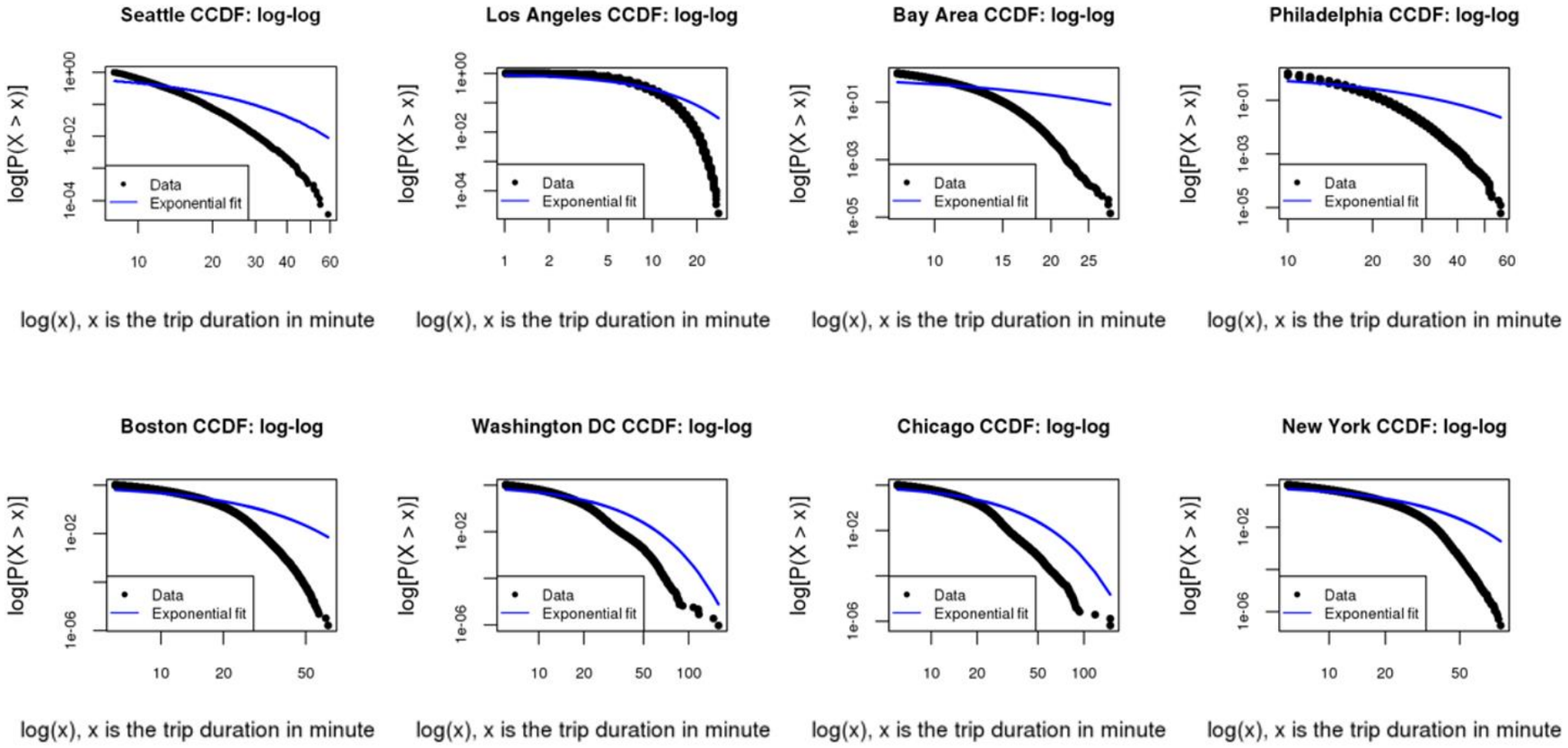


Figure A. 12. Exponential fit to the commuting trip duration

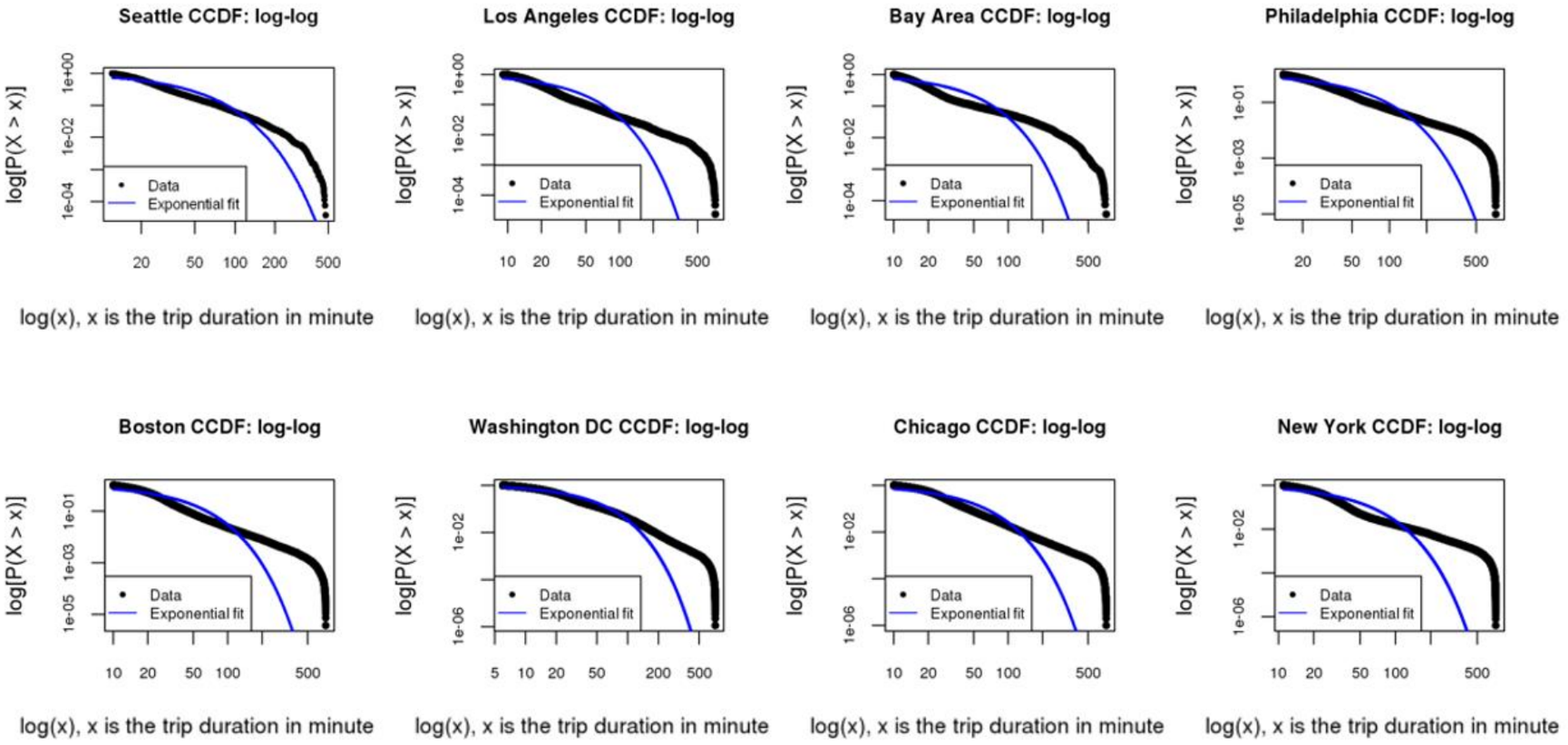


Figure A. 13. Exponential fit to the touristic trip duration

APPENDIX B. SUPPLEMENTAL INFORMATION FOR CHAPTER 3

B.1 Detailed procedures of BS-EREM

The BS-EREM involves the following variables:

T: Bike share trip data in a city C in the study period (with n trips in total);

T_i (ST_i, $Dist_i$, Dur_i, P_i, $transit_{Oi}$, $transit_{Di}$): Trip information of the ith trip in T, i= 1, 2, … , n;

N (ST_N, $Dist_N$, P_N, $Mode_N$): a set of NHTS trips in the same city C;

where

ST: trip start time interval, indicates the time slots (out of the 6 time intervals) that the trip start time lies in;

Dur: duration;

$Dist$: distance;

$Transit_O$: public transit accessibility around bike share trip origin (True/False);

$Transit_D$: public transit accessibility around bike share trip destination (True/False);

P: trip purpose (Commuting/Leisure);

$Mode$: transportation mode of the NHTS trip (Walking/Vehicle);

$Speed_i$: The speed of the ith bike share trip;

$SpeedThreshold$: The speed threshold to determine the trip purpose (commuting/leisure) of a bike share trip (in this study, the speed threshold is the average speed of all bike share trips T);

N_{subset}: A subset of the NHTS trips that meets a certain condition;

$N_{subset,\ Mode\ =\ M}$: NHTS trips in N_{subset} with $Mode_N = M$;

E_M: Emission factor for transportation mode M, where $M \in$ {Biking, PublicTransit, Walking, Vehicle};

S_M: Mode share of transportation mode M in N_{subset}, where $M \in$ {PublicTransit, Walking, Vehicle};

$ReplacedMode_i$: Replaced mode of the ith bike share trip in the simulation;

R: A random number generated in the simulation process ($0 \leq R \leq 1$);

ER_i: Emission reduction of the ith bike share trip in the simulation;

ER_{Total}: Emission reduction of all bike share trip T;

The algorithm below describes the details of BS-EREM in one round of simulation:

Algorithm for bike share trip mode substitution simulation

/* First read in the data and parameters. */
READ NHTS trips N
READ bike share trips T

READ *SpeedThreshold*

READ emission factors E_{Biking}, $E_{PublicTransit}$, $E_{Walking}$, $E_{Vehicle}$

/* Simulate the mode replacement for each trip and determine the emission reduction. */

FOR each trip T_i (i= 1, 2, … , n)

COMPUTE $Speed_i = Dist_i/Dur_i$

IF $Speed_i \geq$ SpeedThreshold **THEN**

P_i = Commuting

ELSE

P_i = Leisure

END

SUBSET NHTS trip data N that trips in N_{subset} have $ST_N = ST_i$, int($Dist_N$) = int($Dist_i$), and $P_N = P_i$

/* The NHTS trip subset N_{subset} have identical trip information (time interval, distance, and purpose with bike share trip T_i; this study uses a 1-mile distance interval to subset NHTS trips. */

SUBSET $N_{subset,\ Mode\ =\ M}$ from N_{subset} where all trips in $N_{subset,\ Mode\ =\ M}$ have $Mode_N = M|$, for M in {PublicTransit, Walking, Vehicle}

COMPUTE the transportation mode share: $S_M = |N_{subset,\ Mode\ =\ M}|/\ |N_{subset}|$, for M in {PublicTransit, Walking, Vehicle}

/* $|N_{subset,\ Mode\ =\ M}|$ and $|N_{subset}|$ are the number of the trips in dataset $N_{subset,\ Mode\ =\ M}$ and N_{subset}, respectively.*/

IF $transit_{Oi}$= False **OR** $transit_{Di}$ = False **THEN**

$S_{PublicTransit} = 0$

$S_M = |N_{subset,\ Mode\ =\ M}|/\ (|N_{subset}|-|N_{subset,\ Mode\ =\ PublicTransit}|/)$, for M in {Walking, Vehicle}

/* When there is no public transit facilities around trip origin or destination, adjust the mode share of public transit to be zero so the bike share trip cannot replace a public transit trip.*/

END

GENERATE a random number R ($0 \leq R \leq 1$)

IF $R \leq S_{PublicTransit}$ **THEN**

$ReplacedMode_i$ = PublicTransit

ELSE IF $S_{PublicTransit} \leq R \leq S_{PublicTransit} + S_{walking}$ **THEN**

$ReplacedMode_i$ = Walking

ELSE

$ReplacedMode_i$ = Vehicle

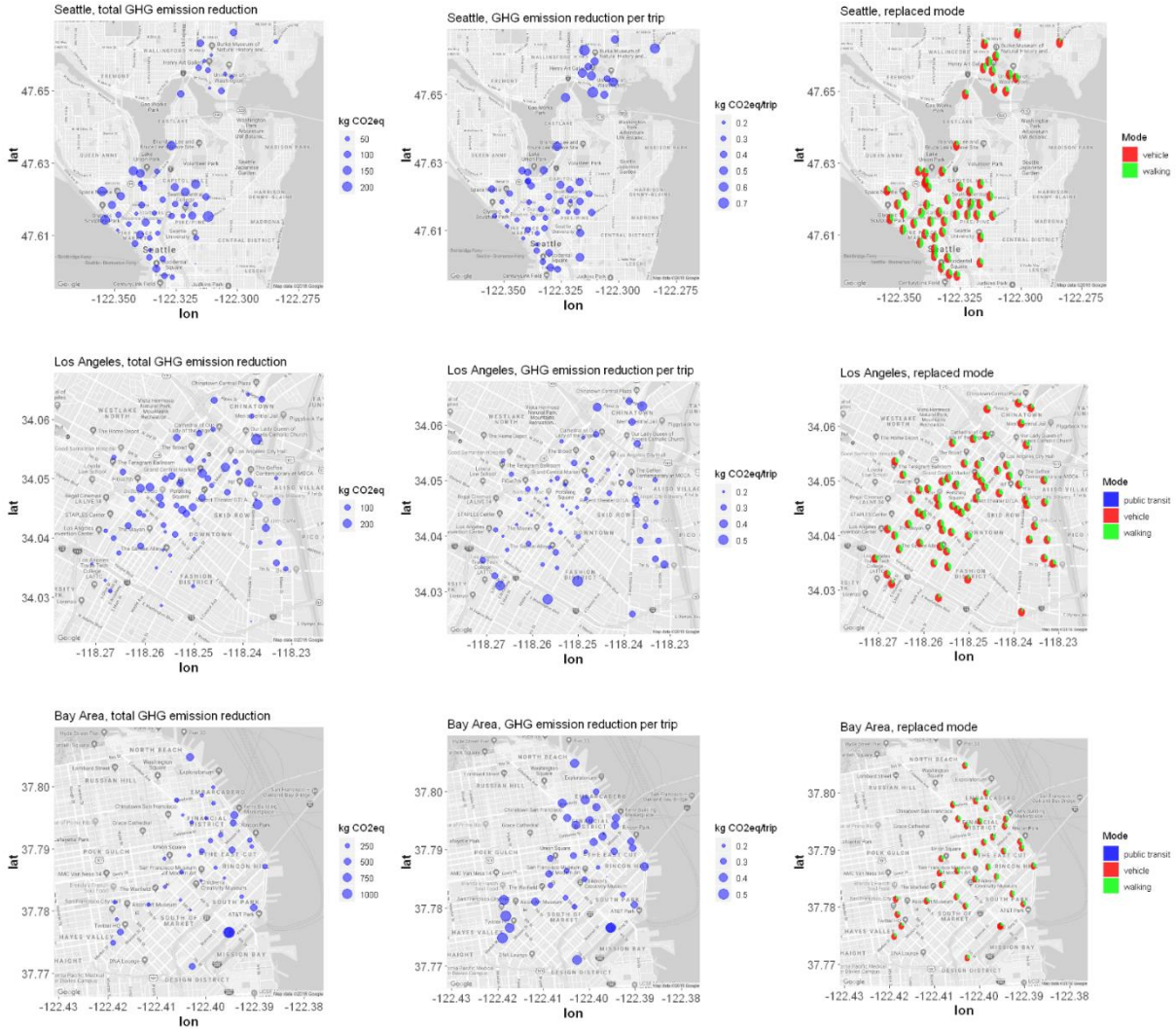
END

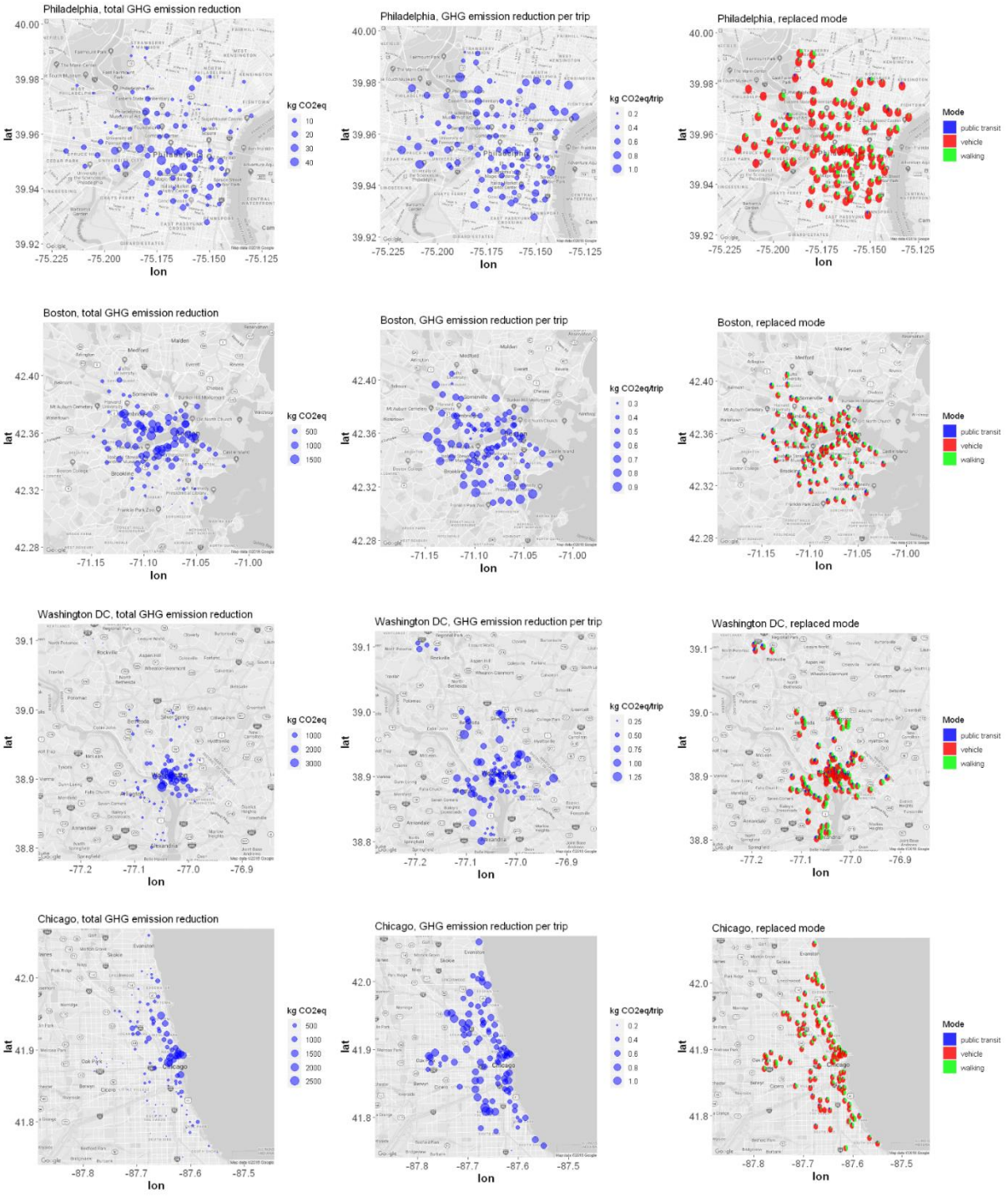
COMPUTE $ER_i = (E_{ReplacedModei} - E_{Biking}) \times Dist_i$ /* Calculate emission reduction of trip T_i */

END

COMPUTE $ER_{total} = \sum_{i=1}^{n} ER_i$, i= 1, 2, … , n /* Calculate the total emission reduction of all trips T. */

B.2 Spatial distribution of the GHG emission reduction for all the eight cities




Philadelphia, total GHG emission reduction
Philadelphia, GHG emission reduction per trip
Philadelphia, replaced mode
Boston, total GHG emission reduction
Boston, GHG emission reduction per trip
Boston, replaced mode
Washington DC, total GHG emission reduction
Washington DC, GHG emission reduction per trip
Washington DC, replaced mode
Chicago, total GHG emission reduction
Chicago, GHG emission reduction per trip
Chicago, replaced mode
lat
lon
kg CO2eq
kg CO2eq/trip
Mode
public transit
vehicle
walking

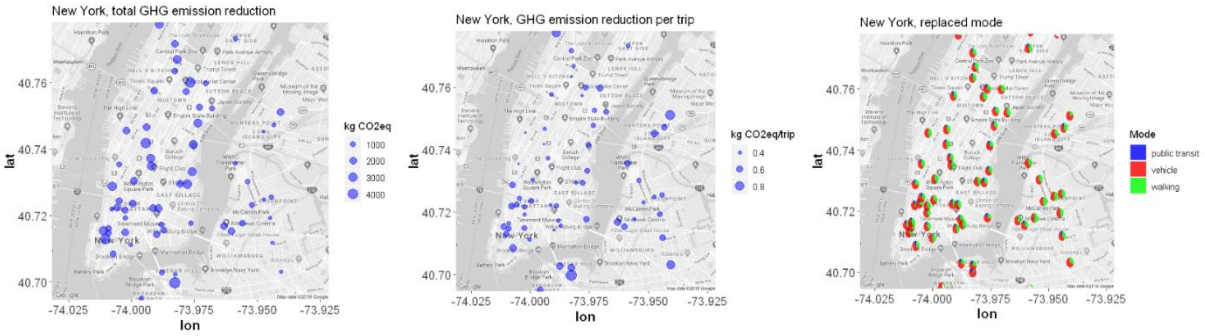


Figure B. 1. Spatial distribution of the emission reduction in different cities (100 stations are randomly selected and visualized in the four larger systems to avoid overlap in visualization

APPENDIX C. SUPPLEMENTAL INFORMATION FOR CHAPTER 4

C.1 Examples of noises in the demand time series

This section shows two examples of the noises in the demand time series, which may lead to very large growth rate. In Figure C. 1, the daily average bike withdrawals at station 16 stayed at a relatively stable level in the high-demand months (May – September) from 2014 to 2019. However, in some low-demand months such as January and February, there was an irregular peak in year 2017, which obviously differs from the demand change pattern in other months. The February demand growth rate from 2016 to 2017 is 2.1, which will bias the results if only analyzing the demand changes for station 16. Similarly, the station 99 in Figure C. 2 exhibited a slight demand decline in most of the months except January and February. The noises in January and February could also influence the analysis of demand change patterns. There are many external factors that could lead to such noises, such as special events and price promotions (e.g., free rides). In this study, the demand of stations are aggregated by the band distance, which can reduce the impact of such noises. Because most of the noises occur when the demand of a station is low, such low demand noises only impact the aggregated demand to a small extent.

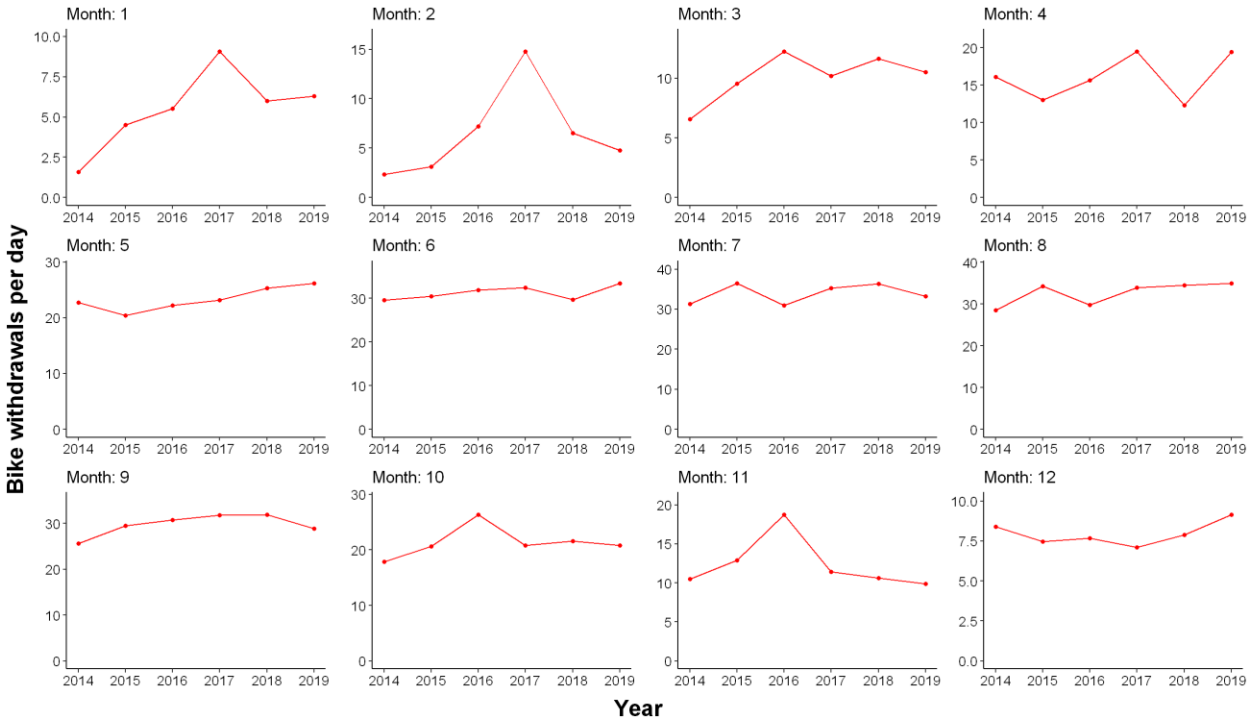


Figure C. 1. Time series of the demands at station 16

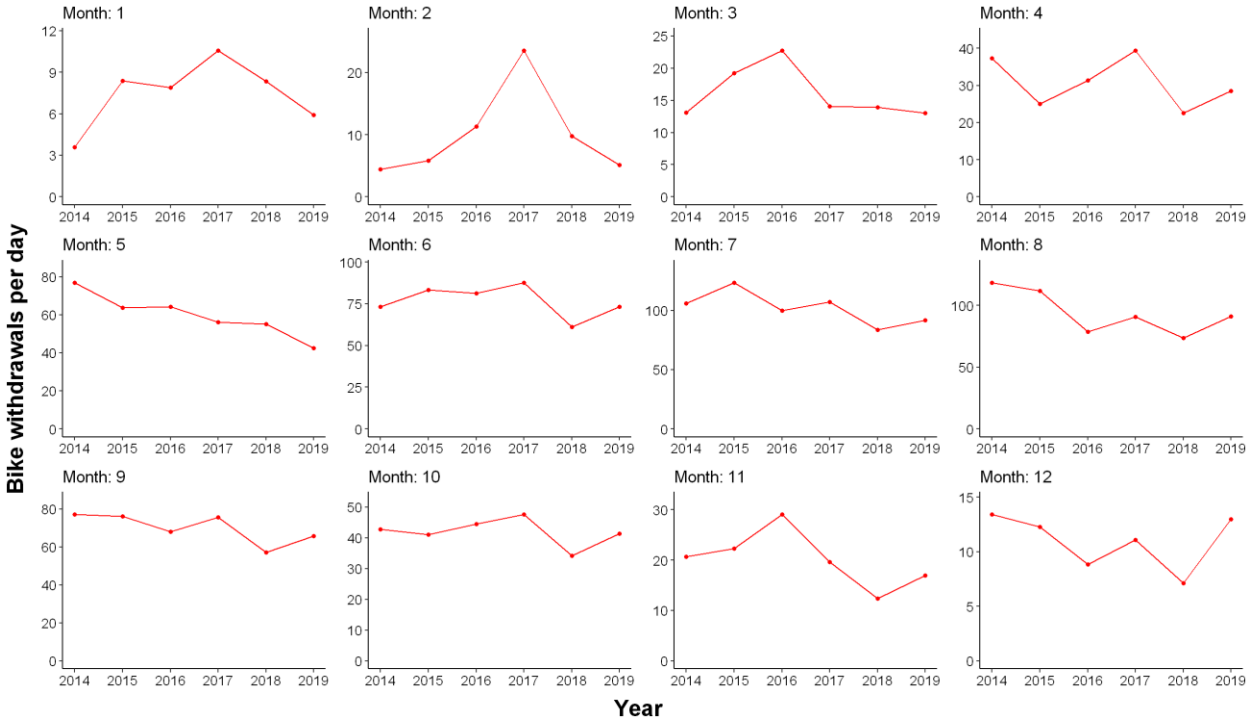


Figure C. 2. Time series of the demands at station 99

C.2 Comparing the results of the piecewise regression model and the linear regression model

This section compares the fitted results of the piecewise regression model and a simple linear regression model. The simple linear regression model has the form in Eq. (C1)

$$E[G \mid D] = \beta_0' + \beta_1' D \tag{C1}$$

where β_0' and β_1' are other parameters to be estimated.

Figure C. 1 shows the fitted results of the simple linear regression model. Compared with the fitted piecewise regression in Figure 4.4, the simple linear regression model fails to capture the trend that the growth rate *G* first increases as the band distance *D* increases when *D* is smaller than the change point of 0.3 miles. The fitted parameters and model fit statistics of the simple linear regression are shown in Table C. 1. The R^2 (0.317) and adjusted R^2 (0.301) of the simple linear regression model are both smaller than the R^2 (0.624) and adjusted R^2 (0.596) of the piecewise regression model, which indicates that the piecewise regression model fits better to the data than the simple linear regression model.

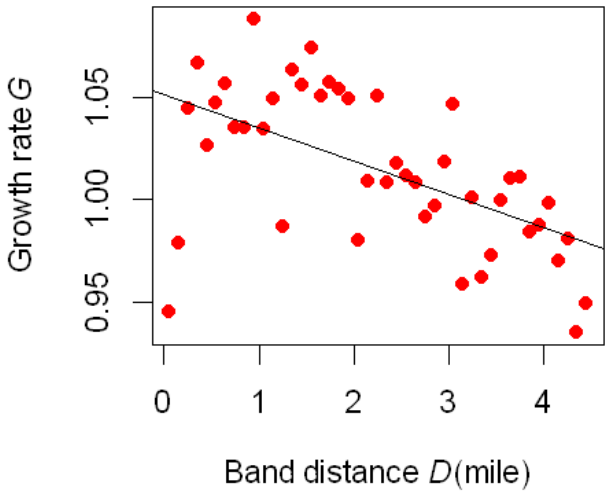


Figure C. 3. Fitted simple linear regression model

Table C. 1. Fitted parameters and the model fit statistics of the simple linear regression model

Parameter	Estimation
β_0'	1.051*** (0.009)
β_1'	-0.016*** (0.004)
Model fit statistics of simple linear regression	
R^2	0.317
Adjusted R^2	0.301
Model fit statistics of piecewise regression	
R^2	0.624
Adjusted R^2	0.596

*Note: * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001;*

the standard errors are shown in parentheses.

C.3 Hyper-parameter tuning for the demand prediction models

For each trained XGBoost model, the hyper-parameters are tuned using grid search (Syarif et al., 2016) with 10-fold cross validation. Different combinations of the hyper-parameters in Table C. 2 are iterated, and the set of hyper-parameters that yields the best cross validation performance will be used for model training. Some examples of the selected hyper-parameters are shown in Table C. 3 to C. 5.

Table C. 2. Tuned hyper-parameters and the corresponding candidate values

Tuned hyper-parameter	Candidate values
Number of boosting rounds (Boosting rounds)	500, 800
Maximum tree depth (Max depth)	6, 8
Boosting learning rate (Learning rate)	0.05, 0.01
Minimum loss reduction required to make a further split on a leaf node (Min loss reduction)	0.2, 0.4
Subsample ratio of the training data records (Data subsample ratio)	0.7, 0.8, 0.9
Subsample ratio of columns for any level (Column subsample ratio)	0.7, 0.8, 0.9

Table C. 3. Selected hyper-parameters for “GEO” model

Year	Month	Column subsample ratio	Min loss reduction	Learning rate	Max depth	Boosting rounds	Data subsample ratio
2014	1	0.7	0.2	0.01	6	500	0.7
	2	0.7	0.2	0.01	6	500	0.7
	3	0.7	0.2	0.01	6	500	0.7
	4	0.7	0.4	0.01	6	500	0.7
	5	0.7	0.4	0.01	6	500	0.7
	6	0.7	0.2	0.01	6	500	0.7
	7	0.7	0.2	0.01	8	500	0.7
	8	0.7	0.4	0.01	8	500	0.7
	9	0.7	0.2	0.01	6	500	0.7
	10	0.7	0.4	0.01	6	500	0.7
	11	0.7	0.4	0.01	6	500	0.7
	12	0.7	0.2	0.01	6	500	0.7
2015	1	0.7	0.4	0.01	6	500	0.7
	2	0.7	0.4	0.01	6	500	0.7
	3	0.7	0.2	0.01	6	500	0.7
	4	0.7	0.4	0.01	8	500	0.7
	5	0.7	0.2	0.01	6	500	0.7
	6	0.7	0.2	0.01	6	500	0.7
	7	0.7	0.4	0.01	6	500	0.7
	8	0.7	0.4	0.01	6	500	0.7
	9	0.7	0.4	0.01	6	500	0.7
	10	0.7	0.2	0.01	6	500	0.7
	11	0.7	0.2	0.01	6	500	0.7
	12	0.7	0.2	0.01	6	500	0.7
2016	1	0.7	0.2	0.01	6	500	0.7
	2	0.7	0.4	0.01	6	500	0.7
	3	0.7	0.4	0.01	6	500	0.7
	4	0.7	0.4	0.01	6	500	0.7
	5	0.7	0.2	0.01	6	500	0.7
	6	0.7	0.4	0.01	6	500	0.7
	7	0.7	0.4	0.01	8	500	0.7
	8	0.7	0.2	0.01	8	500	0.7
	9	0.7	0.2	0.01	8	500	0.7
	10	0.7	0.4	0.01	8	500	0.7
	11	0.7	0.4	0.01	8	500	0.7
	12	0.7	0.2	0.01	6	500	0.8
2017	1	0.7	0.4	0.01	6	500	0.7

	2	0.7	0.4	0.01	6	500	0.7
	3	0.7	0.4	0.01	6	500	0.7
	4	0.7	0.2	0.01	8	500	0.7
	5	0.7	0.2	0.01	8	500	0.7
	6	0.7	0.4	0.01	8	500	0.7
	7	0.7	0.2	0.01	8	500	0.7
	8	0.7	0.2	0.01	8	500	0.7
	9	0.7	0.4	0.01	8	500	0.7
	10	0.7	0.4	0.01	8	500	0.7
	11	0.7	0.2	0.01	8	500	0.7
	12	0.7	0.2	0.01	6	500	0.7
2018	1	0.7	0.2	0.01	6	500	0.7
	2	0.7	0.4	0.01	6	500	0.7
	3	0.7	0.2	0.01	6	500	0.8
	4	0.7	0.4	0.01	8	500	0.7
	5	0.7	0.2	0.01	8	500	0.7
	6	0.7	0.4	0.01	8	500	0.7
	7	0.7	0.4	0.01	8	500	0.7
	8	0.7	0.4	0.01	8	500	0.7
	9	0.7	0.4	0.01	8	500	0.7
	10	0.7	0.4	0.01	8	500	0.7
	11	0.7	0.2	0.01	8	500	0.7
	12	0.7	0.4	0.01	8	500	0.7

Table C. 4. Selected hyper-parameters for “GEO+DENS” model

Year	Month	Column subsample ratio	Min loss reduction	Learning rate	Max depth	Boosting rounds	Data subsample ratio
2014	1	0.7	0.4	0.01	6	500	0.7
	2	0.7	0.2	0.05	8	500	0.8
	3	0.7	0.2	0.01	8	500	0.7
	4	0.9	0.2	0.01	8	500	0.9
	5	0.9	0.2	0.05	8	500	0.9
	6	0.7	0.4	0.05	6	500	0.8
	7	0.8	0.4	0.05	6	500	0.8
	8	0.8	0.2	0.05	6	500	0.8
	9	0.9	0.4	0.01	8	500	0.9
	10	0.7	0.2	0.05	6	500	0.9
	11	0.7	0.4	0.05	6	500	0.7
	12	0.7	0.4	0.05	8	500	0.8
2015	1	0.7	0.2	0.01	6	500	0.7
	2	0.7	0.4	0.01	8	500	0.7
	3	0.7	0.2	0.05	8	500	0.7
	4	0.9	0.2	0.01	8	500	0.7
	5	0.8	0.2	0.01	8	500	0.7
	6	0.9	0.2	0.01	8	500	0.8
	7	0.8	0.4	0.01	8	500	0.7
	8	0.8	0.4	0.01	8	500	0.7
	9	0.7	0.2	0.01	6	500	0.7
	10	0.9	0.4	0.01	6	500	0.7
	11	0.9	0.2	0.01	6	500	0.7
	12	0.7	0.4	0.01	6	500	0.8
2016	1	0.8	0.4	0.01	6	500	0.7
	2	0.9	0.2	0.01	6	500	0.7
	3	0.7	0.4	0.05	8	500	0.8
	4	0.9	0.4	0.01	6	500	0.7
	5	0.9	0.4	0.05	6	500	0.7
	6	0.9	0.4	0.01	6	500	0.8
	7	0.9	0.4	0.01	8	500	0.7
	8	0.8	0.2	0.01	6	500	0.7
	9	0.9	0.4	0.05	8	500	0.8
	10	0.8	0.2	0.05	8	500	0.7
	11	0.9	0.4	0.05	8	500	0.7
	12	0.8	0.2	0.01	8	500	0.7
2017	1	0.7	0.4	0.05	6	500	0.8

	2	0.9	0.2	0.01	6	500	0.8
	3	0.9	0.2	0.01	8	500	0.8
	4	0.7	0.4	0.05	8	500	0.7
	5	0.9	0.4	0.01	8	500	0.7
	6	0.9	0.4	0.01	8	500	0.7
	7	0.9	0.4	0.01	8	500	0.7
	8	0.9	0.4	0.01	8	500	0.7
	9	0.7	0.4	0.01	6	500	0.7
	10	0.9	0.2	0.01	6	500	0.9
	11	0.7	0.2	0.05	8	500	0.7
	12	0.7	0.2	0.01	6	500	0.7
2018	1	0.7	0.4	0.05	8	500	0.8
	2	0.9	0.2	0.01	6	500	0.9
	3	0.7	0.2	0.01	6	500	0.7
	4	0.7	0.2	0.01	6	500	0.7
	5	0.7	0.4	0.01	8	500	0.7
	6	0.8	0.2	0.01	8	500	0.7
	7	0.7	0.4	0.01	6	500	0.7
	8	0.9	0.4	0.01	6	500	0.7
	9	0.7	0.2	0.01	6	500	0.7
	10	0.9	0.2	0.05	6	500	0.7
	11	0.7	0.2	0.01	8	500	0.7
	12	0.7	0.4	0.05	8	500	0.7

Table C. 5. Selected hyper-parameters for “EXT” model

Year	Month	Column subsample ratio	Min loss reduction	Learning rate	Max depth	Boosting rounds	Data subsample ratio
2014	1	0.7	0.2	0.01	6	500	0.8
	2	0.7	0.4	0.01	6	500	0.7
	3	0.7	0.4	0.01	6	500	0.7
	4	0.7	0.2	0.01	6	500	0.7
	5	0.7	0.2	0.01	8	500	0.7
	6	0.7	0.2	0.01	8	500	0.7
	7	0.7	0.4	0.01	8	500	0.7
	8	0.7	0.4	0.01	6	500	0.7
	9	0.8	0.2	0.01	8	500	0.7
	10	0.7	0.4	0.01	6	500	0.7
	11	0.8	0.4	0.01	6	500	0.7
	12	0.8	0.2	0.01	6	500	0.7
2015	1	0.7	0.2	0.01	6	500	0.7
	2	0.7	0.2	0.05	8	500	0.8
	3	0.8	0.2	0.01	6	500	0.7
	4	0.7	0.2	0.01	6	500	0.7
	5	0.7	0.4	0.01	8	500	0.7
	6	0.7	0.2	0.01	8	500	0.7
	7	0.8	0.4	0.01	8	500	0.8
	8	0.7	0.4	0.01	6	500	0.7
	9	0.7	0.4	0.01	6	500	0.9
	10	0.7	0.2	0.01	6	500	0.7
	11	0.8	0.4	0.01	6	500	0.7
	12	0.7	0.4	0.01	8	500	0.7
2016	1	0.7	0.2	0.01	6	500	0.7
	2	0.7	0.2	0.01	6	500	0.7
	3	0.9	0.2	0.01	6	500	0.7
	4	0.7	0.4	0.01	8	500	0.7
	5	0.7	0.4	0.01	6	500	0.8
	6	0.8	0.4	0.01	8	500	0.7
	7	0.7	0.2	0.01	6	500	0.7
	8	0.7	0.4	0.01	8	500	0.7
	9	0.8	0.4	0.01	8	500	0.7
	10	0.7	0.2	0.01	6	500	0.7
	11	0.7	0.4	0.01	6	500	0.7
	12	0.8	0.2	0.01	6	500	0.7
2017	1	0.9	0.4	0.01	6	500	0.7

	2	0.7	0.4	0.01	6	500	0.7
	3	0.9	0.2	0.01	6	500	0.7
	4	0.7	0.4	0.01	6	500	0.7
	5	0.7	0.2	0.01	6	500	0.7
	6	0.7	0.4	0.01	6	500	0.9
	7	0.8	0.2	0.01	6	500	0.7
	8	0.7	0.2	0.01	8	500	0.8
	9	0.7	0.2	0.01	6	500	0.7
	10	0.8	0.2	0.01	6	500	0.7
	11	0.8	0.4	0.01	6	500	0.7
	12	0.7	0.4	0.01	6	500	0.7
2018	1	0.7	0.4	0.01	6	500	0.7
	2	0.7	0.2	0.05	8	800	0.7
	3	0.8	0.2	0.01	8	500	0.7
	4	0.7	0.2	0.01	6	500	0.7
	5	0.7	0.4	0.01	6	500	0.7
	6	0.7	0.4	0.01	6	500	0.8
	7	0.7	0.2	0.01	8	500	0.7
	8	0.8	0.4	0.01	8	500	0.7
	9	0.7	0.4	0.01	6	500	0.7
	10	0.7	0.4	0.01	8	500	0.7
	11	0.8	0.2	0.01	8	500	0.7
	12	0.7	0.4	0.01	6	500	0.8

APPENDIX D. SUPPLEMENTAL INFORMATION FOR CHAPTER 5

D.1 Computing the biking O-D correlation matrix

As stated in Section 5.3.3.2, in the biking O-D correlation matrix M_t, the element m_t^{ij} is the percentage of trips that end in station *j* given that they start from station *i* in hour *t* of the *T* days. Similar to the estimation of true bike withdrawal rate, the model also needs to consider the impact of the empty/full stations. Therefore, this study first computes the average hourly O-D trip rate $\hat{R}_t^{ij}$ (Eq. D1) for trips that start from station *i* and end at station *j* during time interval *t* over *T* days, where $\hat{w}_{tz}^{ij}$ is the number of observed bike share trips that start from station *i* and end at station *j* during time interval *t* on the z^{th} day, and $\hat{p}_t^{ij}$ is the percentage of time that station *i* has available bikes and station *j* has available docks at the same time during time *t* over *T* days. Then m_t^{ij} can be computed using Eq. D2.

$$\hat{R}_t^{ij} = \frac{1}{T \cdot \hat{p}_t^{ij}} \sum_{z=1}^{T} \hat{w}_{tz}^{ij} \tag{D1}$$

$$m_t^{ij} = \frac{\hat{R}_t^{ij}}{\sum_j \hat{R}_t^{ij}} \tag{D2}$$

This study first simulates the biking origin of a trip and then assigns a destination based on the correlation matrix M_t. However, since the average hourly O-D trip rate $\hat{R}_t^{ij}$ is already estimated in (Eq. D1), another option of generating trips is to directly simulate the trips based on $\hat{R}_t^{ij}$ for O-D pairs, such as the "Filtered Sample Average" method in the study of (Goh et al., 2019). However, the true demand estimation based on the station status may over-estimate the demand. For instance, it could happen that there is indeed no bike withdrawal demand at an empty station for a certain period, in which case the estimated demand rate is larger than the actual demand. If the station status at both the origins and destinations are considered, it may further over-estimate the true demand. As a compromise, this study only considers the origin station statues to estimate the true

demand to reduce the over-estimation. But the impact of using $\hat{R}_t^{ij}$ to directly generate the trips.is also evaluated The total estimated true demand (trips) per day considering both O-D station status is 0.3%, 3.1%, and 5.6% more than the true demand estimated based on only the origin station status in Los Angeles, Philadelphia, and Chicago, respectively.

D.2 Bike pick-up in the user behavior model

In this study, a trip's actual origin is estimated based on the biking origin (primary grid) according to the distance distribution in $WALK_{first}$, to be consistent with the historical data, a trip whose actual origin is one grid away from the primary grid inherently indicates that the user of the trip is at least willing to walk one grid to pick up bikes. For example, for a trip served by a station-based BSS, if the trip's actual origin is estimated to be one grid away from the primary grid, it is known that at least the user is willing to walk to the primary grid to use the system. Therefore, this user should also be willing to walk to other grids that are one-grid away to pick up bikes when the bike availability is not constrained by the stations in the dockless and hybrid systems. This revealed walking tolerance is defined as d_{tor}, which is the distance between the actual origin grid and the corresponding primary grid. If the distance between the actual origin and the nearest bike, d_{bike}, is less than or equal to d_{tor}, the user will pick up the nearest bike and the trip will always be served. However, someone who is at least willing to walk one grid to use the BSS may also be willing to walk two grids. In the case that $d_{bike} > d_{tor}$, the probability of picking up the bike is the conditional probability described in Eq. D3.

$$\begin{aligned} &\mathrm{P}\left(\text{the user will walk } d_{bike} \mid \text{ the user is at least willing to walk } d_{tor}\right) \\ &= \frac{\mathrm{CCDF}(d_{bike})}{\mathrm{CCDF}(d_{tor})} \end{aligned} \tag{D3}$$

where CCDF is the Complementary Cumulative Distribution Function of the walking distance distribution in $WALK_{first}$. Figure D. 1 shows one example of such bike pick-up process.

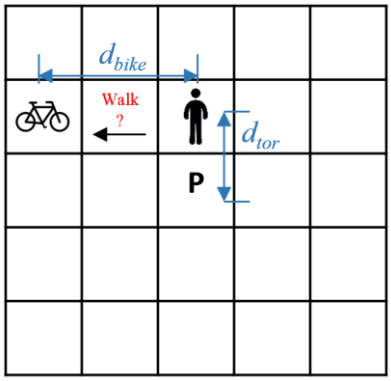


Figure D. 1. One example of the bike pick-up process: the user of this trip has d_{tor} = one-grid because the starting location is one grid away from the primary grid (the grid with a "P" sign). The nearest available bike is two grids away from the user, i.e., d_{bike} = two-grids. In this case, the probability that the user will walk to pick up this bike equals to CCDF(d_{bike})/CCDF(d_{tor})

D.3 Comparing rebalance methods

This section compares the performance of rebalance method proposed by O'Mahony and Shmoys (2015) (referred as "literature method" in the following part) with the proposed method. The literature method is formulated as below in Eq. D4 to D9 to allocate the N bikes to all the stations. The notations have the same meanings as those in the main text. When the system has sufficient bikes, the literature method tends to allocate the bike inventory level L_i at $\frac{C_i - \min_\tau f_\tau^i - \max_\tau f_\tau^i}{2}$ for station i. In this case, the capacity C_i not only plays a role as a limit, but also affects the assigned level L_i. However, some stations may have larger capacity than its actual demands. Assigning more bikes to these large-capacity but low-usage stations will reduce the bikes allocated to stations with smaller capacity but high usage, causing more unserved trips. Therefore, in this study, the working capacity C_i' is introduced.

$$\min \sum_i D_i \tag{D4}$$

$$s.t. \quad \sum_i L_i = N \tag{D5}$$

$$\forall i \quad D_i \geq L_i + \min_\tau f_\tau^i \tag{D6}$$

$$\forall i \quad D_i \geq C_i - L_i - \max_\tau f_\tau^i \tag{D7}$$

$$\forall i \qquad L_i \in \{0,\ 1, \ldots, C_i\} \tag{D8}$$

$$\forall i \qquad D_i \geq 0 \tag{D9}$$

As shown in Figure D. 2, the proposed method has significantly reduced the percentage of unserved trips in the station-based BSSs in all three cities compared with the literature method. Although this method increased full-dock trips, the total percentage of affected trips is still lower than the literature method. It is notable that a lower percentage of unserved trips also means that more trips were started, which may indirectly increase the probability to encounter full-dock issues.

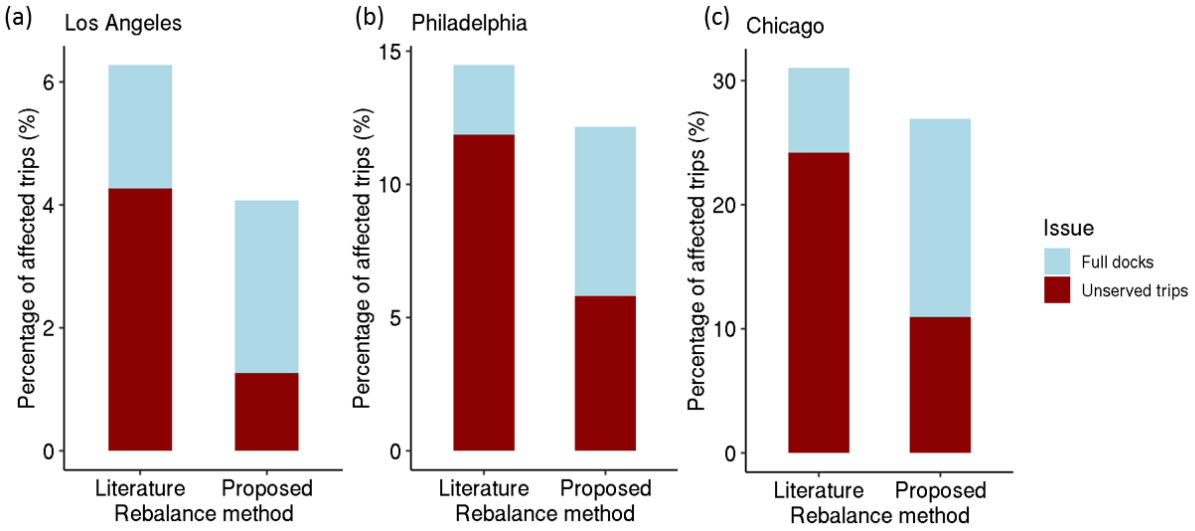


Figure D. 2. Percentage of affected trips from simulations using the literature rebalance method (O'Mahony & Shmoys, 2015) and the proposed rebalance method.

D.4 Estimation of the rebalance mileage

In this study, the rebalance workload is estimated by computing the mileage to travel through the grids that require rebalance. A traveling salesman problem (TSP) is solved to traverse all rebalancing stations using the method arbitrary insertion (Rosenkrantz et al., 2009), which yields the near-optimal shortest tour length (Hahsler & Hornik, 2007). A two-edge exchange improvement procedure (Hahsler & Hornik, 2007) was also applied to further improve the route if possible. It is assumed that there is only one rebalance vehicle. Similar to the study by Goh et al.

(Goh et al., 2019), the model focuses on estimating the rebalance needs in the BSS and ignores the routing time and time to load and unload bikes. Additionally, based on the communication with BSS operators, a station will not be rebalanced if only one or two bikes need to be added/removed. Therefore, in the simulation, if the rebalance need at a station or a primary grid catchment is too small (e.g., the absolute difference between the current inventory level and the solved target level is less than or equal to 2 bikes), the rebalance will be delayed to later when the current inventory level further deviate from the target level. Because of having this flexibility, sometimes the number of move-out bikes does not equal to the move-in bikes across the entire system. Therefore, the depot is used to provide a buffer capacity, supplying or storing a small amount of bikes to adjust for the differences.

D.5 Spatial distribution of the bike withdrawal demand in Philadelphia and Chicago

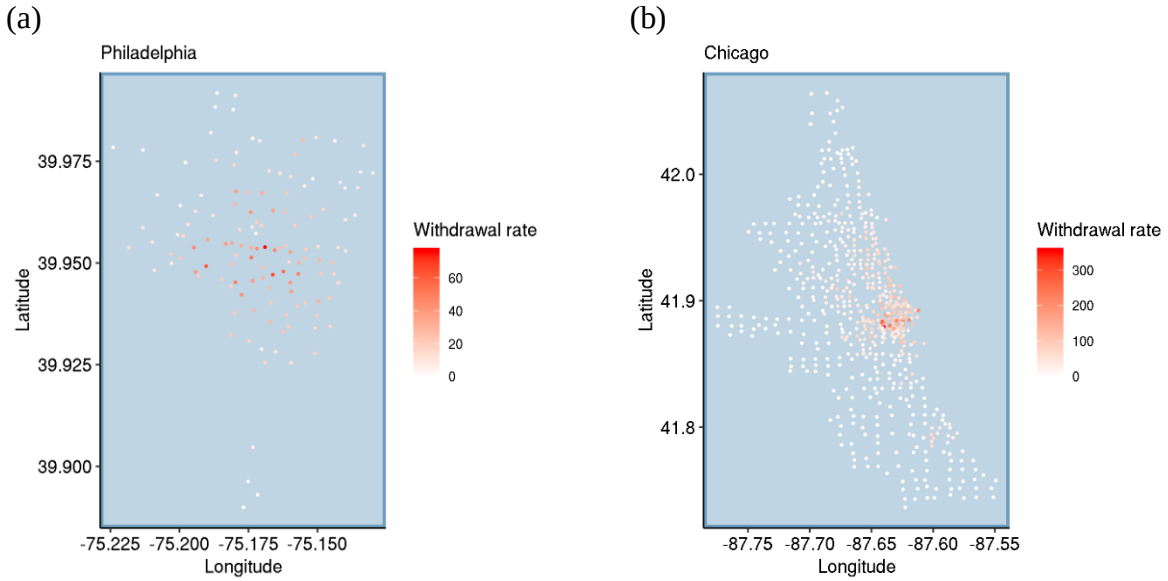


Figure D. 3. Heat map of the bike withdrawal rate per day for the stations in (a) Philadelphia, and (b) Chicago

D.6 Detailed results of access time

Table D. 1. Average excess time (seconds) in different stages per served trips

City	System type	Start-walk	End-walk	End-bike	Total
Los Angeles	Dockless	36.8	0.0	0.0	36.8
	Hybrid(25%)	38.6	8.1	0.0	46.7
	Hybrid(50%)	39.8	14.9	0.0	54.7
	Hybrid(75%)	41.5	22.3	0.0	63.8
	Station	45.6	62.6	16.1	124.3
Philadelphia	Dockless	36.4	0.0	0.0	36.4
	Hybrid(25%)	38.4	8.9	0.0	47.3
	Hybrid(50%)	40.0	17.4	0.0	57.4
	Hybrid(75%)	42.3	25.3	0.0	67.6
	Station	47.1	53.3	9.2	109.6
Chicago	Dockless	36.3	0.0	0.0	36.3
	Hybrid(25%)	38.0	7.8	0.0	45.9
	Hybrid(50%)	39.9	15.2	0.0	55.1
	Hybrid(75%)	41.7	22.5	0.0	64.2
	Station	46.4	71.2	25.5	143.1

D.7 Detailed analysis of bike rebalance

This section provides more detailed analysis on the bike rebalance in Chicago. In the main manuscript, it is shown that the station-based BSS in Chicago requires more rebalance mileages than the corresponding hybrid BSS. The examples of rebalance routes in Figure D. 4 show that there are some outskirt grids in Chicago that are visited by the rebalance vehicle in the station-based BSSs but not visited in the corresponding dockless/hybrid BSSs. Visiting these outskirt grids requires more mileage than only circling around in the downtown area, which makes the total rebalance mileage of station-based BSS higher than the corresponding hybrid BSS in Chicago. Figure D. 5 shows the inventory changes of two such grids. In both grids, the assigned initial inventory level in the station-based BSS is higher than the dockless BSS. Grid A has an initial inventory of three bikes in the station-based BSS, which became empty at the end of the day and thus required rebalance. However, grid A was only assigned one bike in the dockless BSS and also became empty at the end of the day (with three unserved trips). Because the difference between the end-of-day inventory and the rebalance target level is only one bike, less than the 2-bike threshold to trigger a rebalance request, rebalance is not conducted at Grid A in the dockless BSS. Grid B in the station-based BSS had an initial inventory of two bikes at the beginning of the day and the inventory changed to nine at the end of the day, which required rebalance. However, in the corresponding dockless BSS, the initial and end-of-day bike inventory are 0 and 1, respectively, and thus Grid B does not need a rebalance. It is also notable that Grid B received less returned bikes in the dockless BSS than in the station-based BSS throughout the day. This is because that the dockless/hybrid BSSs can allocate more bikes to the grids with high bike withdrawal demands, while the station-based BSSs have to consider the capacity limits and the trade-off between the availability of both bikes and docks. As a result, in dockless/hybrid BSS, more bikes are located at the high demand downtown area, while outskirts grids are assigned a very small number of bikes (Figure D. 6). In contrast, station-based BSSs allocated more bikes to the outskirt areas than the dockless/hybrid BSSs, which increased the served trips in these areas. Because bike share trips are mostly in short distance (median distance is around 1.5 miles (Kou & Cai, 2019)), when the local bike availability around Grid B is low in a dockless BSS, the number of bikes returned to Grid B would also be reduced due to unserved trips (e.g., the unserved trips in grid A).

The reason that such situations are only observed in Chicago but not in other two cities is that: in Chicago $-\sum_i (\min_\tau f_\tau^i \mid \min_\tau f_\tau^i < 0)$ = 4,030 while the total bike supply N = 3,930, so in the dockless/hybrid BSSs, the rebalance algorithm can only allocate bikes to grids with $\min_\tau f_\tau^i < 0$, and leave other grids empty. While in Los Angeles and Philadelphia, the values of $-\sum_i (\min_\tau f_\tau^i \mid \min_\tau f_\tau^i < 0)$ and N are 201 and 986, and 560 and 1,140, respectively. Therefore, in these two cities, there are still many bikes left after allocating the needed bikes to those grids with $\min_\tau f_\tau^i < 0$. So more bikes can be allocated to other grids, including those outskirt grids.

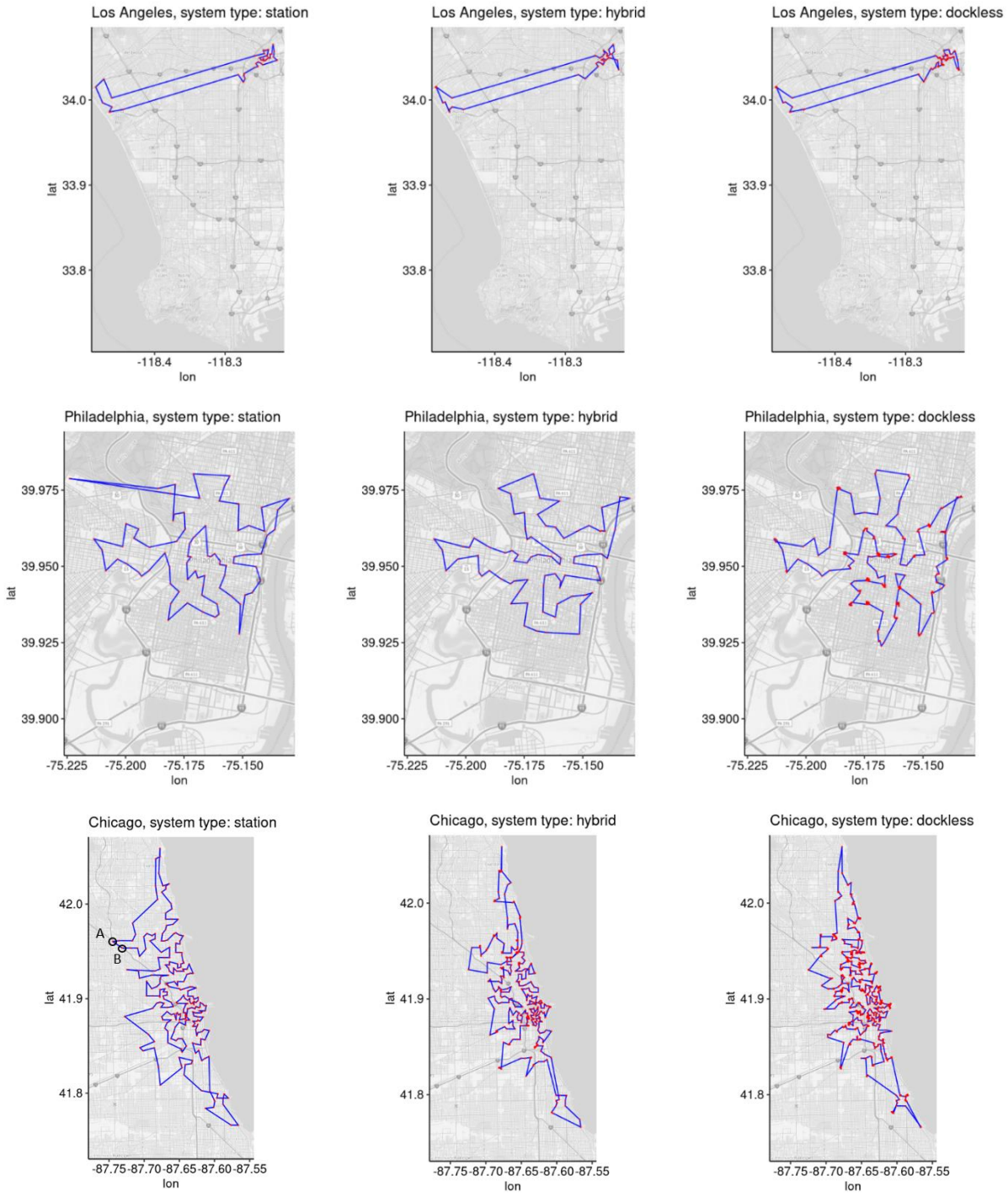


Figure D. 4. Examples of rebalance routes in one round of simulations (blue lines are the routes and each red dot represents one stop that the vehicle stops to collect or drop off bikes so there could be multiple stops in one grid in dockless/hybrid BSSs; the background map covers the whole service areas of the systems but grids not on the rebalance route are not plotted)

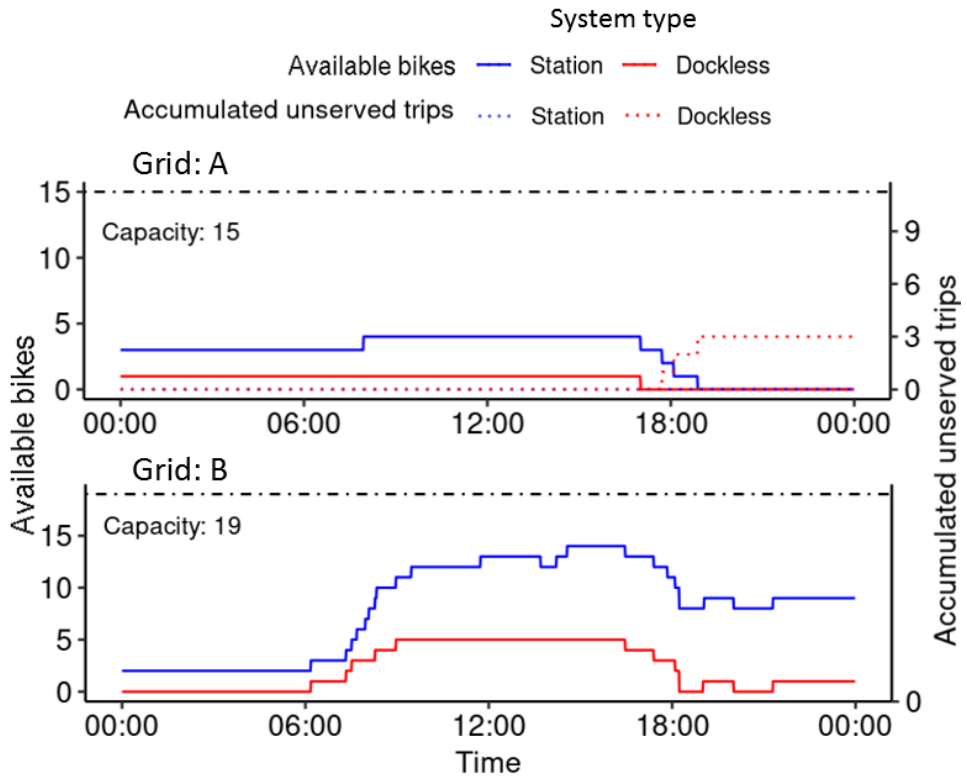
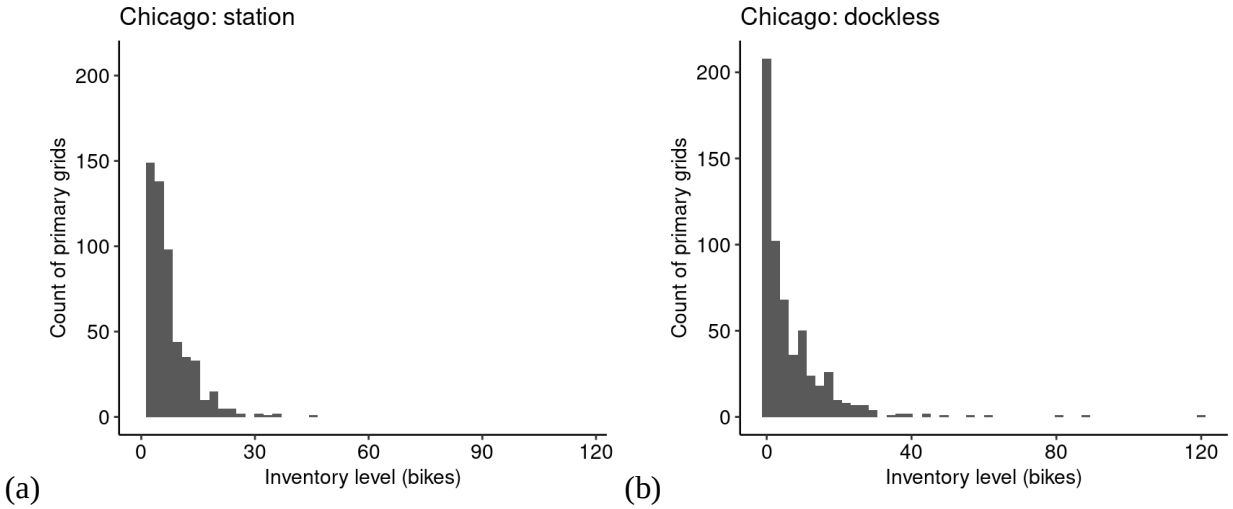


Figure D. 5. The inventory and the accumulated unserved trips dynamic changes of two grids in Chicago (the locations of the two grids are shown in the example rebalance routing in the station-based system in Chicago in Figure D. 4)



(a) (b)

Figure D. 6. Distribution of the assigned initial inventory in Chicago: (a) station-based system, (b) dockless system

REFERENCES

Albiński, S., Fontaine, P., & Minner, S. (2018). Performance analysis of a hybrid bike sharing system: A service-level-based approach under censored demand observations. *Transportation Research Part E: Logistics and Transportation Review*, *116*(November 2017), 59–69. https://doi.org/10.1016/j.tre.2018.05.011

Alter, L. (2008). Bikes Faster than Subway or Car, Seven Years Running. Retrieved November 18, 2017, from TreeHugger website: https://www.treehugger.com/bikes/bikes-faster-than-subway-or-car-seven-years-running.html

Alvarez-Valdes, R., Belenguer, J. M., Benavent, E., Bermudez, J. D., Muñoz, F., Vercher, E., & Verdejo, F. (2016). Optimizing the level of service quality of a bike-sharing system. *Omega (United Kingdom)*, *62*, 163–175. https://doi.org/10.1016/j.omega.2015.09.007

American Community Survey. (2017). ACS Demographic and Housing Estimates - 5-Year Estimates Data Profiles. Retrieved July 1, 2019, from https://data.census.gov/cedsci/table?g=0400000US17_0500000US17031&d=ACS 5-Year Estimates Data Profiles&tid=ACSDP5Y2017.DP05

Anderson, N. A. (2015). *Portland Bicycle Share Health Impact Assessment*.

Atkinson, R. P. D., Rhodes, C. J., Macdonald, D. W., & Anderson, R. M. (2002). Scale-free dynamics in the movement patterns of jackals. *Oikos*, *98*(1), 134–140. https://doi.org/10.1034/j.1600-0706.2002.980114.x

Bazzani, A., Giorgini, B., Rambaldi, S., Gallotti, R., & Giovannini, L. (2010). Statistical laws in urban mobility from microscopic GPS data in the area of Florence. *Journal of Statistical Mechanics: Theory and Experiment*, *2010*(05), P05001. https://doi.org/10.1088/1742-5468/2010/05/P05001

Borecki, N., Rawls, B., Buck, D., Reyes, P., & Al., E. (2012). *Virginia Tech Capital Bikeshare Study*. (January).

Boston, G. (2014). *GHG Change 2012-2013*.

Brinkmann, J., Ulmer, M. W., & Mattfeld, D. C. (2015). Short-term strategies for stochastic inventory routing in bike sharing systems. *Transportation Research Procedia*, *10*(July), 364–373. https://doi.org/10.1016/j.trpro.2015.09.086

Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. *Nature*, *439*(7075), 462–465. https://doi.org/10.1038/nature04292

Buck, D., & Buehler, R. (2012). Bike lanes and other determinants of capital bikeshare trips. *91st Transportation Research Board Annual Meeting*, 703–706.

Buehler, R., & Hamre, A. (2015). Business and Bikeshare User Perceptions of the Economic Benefits of Capital Bikeshare. *Transportation Research Record: Journal of the Transportation Research Board*, *2520*, 100–111. https://doi.org/10.3141/2520-12

Bullock, C., Brereton, F., & Bailey, S. (2017). The economic contribution of public bike-share to the sustainability and efficient functioning of cities. *Sustainable Cities and Society*, *28*, 76–87. https://doi.org/10.1016/j.scs.2016.08.024

Caggiani, L., Camporeale, R., Ottomanelli, M., & Szeto, W. Y. (2018). A modeling framework for the dynamic management of free-floating bike-sharing systems. *Transportation Research Part C: Emerging Technologies*, *87*(April 2017), 159–182. https://doi.org/10.1016/j.trc.2018.01.001

Caggiani, L., & Ottomanelli, M. (2012). A Modular Soft Computing based Method for Vehicles Repositioning in Bike-sharing Systems. *Procedia - Social and Behavioral Sciences*, *54*, 675–684. https://doi.org/10.1016/j.sbspro.2012.09.785

Caggiani, L., & Ottomanelli, M. (2013). A Dynamic Simulation based Model for Optimal Fleet Repositioning in Bike-sharing Systems. *Procedia - Social and Behavioral Sciences*, *87*, 203–210. https://doi.org/10.1016/j.sbspro.2013.10.604

Cai, H., Zhan, X., Zhu, J., Jia, X., Chiu, A. S. F., & Xu, M. (2016). Understanding taxi travel patterns. *Physica A: Statistical Mechanics and Its Applications*, *457*, 590–597. https://doi.org/10.1016/j.physa.2016.03.047

California Department of Transportation. (2013). *2010 --2012 California Household Travel Survey Final Report*. 147. Retrieved from https://catalog.data.gov/dataset/california-household-transportation-survey

Cazabet, R., Borgnat, P., & Jensen, P. (2017). Using degree constrained gravity null-models to understand the structure of journeys' networks in bicycle sharing systems. *ESANN 2017 - Proceedings, 25th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 405–410.

Chai, D., Wang, L., & Yang, Q. (2018). Bike flow prediction with multi-graph convolutional networks. *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, 397–400. https://doi.org/10.1145/3274895.3274896

Chemla, D., Pradeau, T., Calvo, R. W., Chemla, D., Pradeau, T., & Calvo, R. W. (2013). *Self-service bike sharing systems : simulation , repositioning , pricing*.

Chen, L., Jakubowicz, J., Zhang, D., Wang, L., Yang, D., Ma, X., … Nguyen, T.-M.-T. (2016a). *Dynamic cluster-based over-demand prediction in bike sharing systems*. 841–852. https://doi.org/10.1145/2971648.2971652

Chen, L., Zhang, D., Wang, L., Yang, D., Ma, X., Li, S., … Jakubowicz, J. (2016b). *Dynamic Cluster-Based Over-Demand Prediction in Bike Sharing Systems*.

Chen, M., Wang, D., Sun, Y., Waygood, E. O. D., & Yang, W. (2020a). A comparison of users' characteristics between station-based bikesharing system and free-floating bikesharing system: case study in Hangzhou, China. *Transportation*, *47*(2), 689–704. https://doi.org/10.1007/s11116-018-9910-7

Chen, Q., Liu, M., & Liu, X. (2018). Bike Fleet Allocation Models for Repositioning in Bike-Sharing Systems. *IEEE Intelligent Transportation Systems Magazine*, *10*(1), 19–29. https://doi.org/10.1109/MITS.2017.2776129

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-Augu*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, W., Liu, Q., Zhang, C., Mi, Z., Zhu, D., & Liu, G. (2020b). Characterizing the stocks, flows, and carbon impact of dockless sharing bikes in China. *Resources, Conservation and Recycling*, *162*(April), 105038. https://doi.org/10.1016/j.resconrec.2020.105038

Cheng, G., Guo, Y., Chen, Y., & Qin, Y. (2019). Designating City-Wide Collaborative Geofence Sites for Renting and Returning Dock-Less Shared Bikes. *IEEE Access*, *7*, 35596–35605. https://doi.org/10.1109/ACCESS.2019.2903521

Chiariotti, F., Pielli, C., Zanella, A., & Zorzi, M. (2018). A dynamic approach to rebalancing bike-sharing systems. *Sensors (Switzerland)*, *18*(2), 1–22. https://doi.org/10.3390/s18020512

City of Philadelphia. (2019). Philadelphia Dockless Bike Share Pilot. Retrieved September 25, 2019, from http://www.phillyotis.com/portfolio-item/dockless-bike-share-pilot/

Cohen, A., & Shaheen, S. (2018). *Planning for Shared Mobility*. https://doi.org/10.7922/G2NV9GDD

Contardo, C., Rousseau, L.-M., & Morency, C. (2012). Balancing a dynamic public bike-sharing system. *The First Conference of EURO Working Group on Vehicle Routing and Logistics Optimization VeRoLog*, 57–58. Retrieved from http://www.verolog.eu/

Croci, E., & Rossi, D. (2014). Optimizing the Position of Bike Sharing Stations. The Milan Case. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2461179

D'Agostino RB, S. M. (1986). *Goodness-of-Fit Techniques* (1st ed.). Dekker.

Datner, S., Raviv, T., Tzur, M., & Chemla, D. (2015). Setting Inventory Levels in a Bike Sharing Network. *Transportation Science*, (March 2016), 1–20. https://doi.org/10.13140/RG.2.1.4903.9444

Dave, S. (2010). *Life Cycle Assessment of Transportation Options for Commuters*. 1–16.

Department of Energy and Environment. (2012). *District of Columbia Greenhouse Gas Inventory Update 2012-2013*. Retrieved from https://doee.dc.gov/sites/default/files/dc/sites/ddoe/service_content/attachments/2013 Greenhouse Gas Inventory Update_web.pdf

Dickey, M. R. (2019). SF will allow additional dockless bike-share operators. Retrieved July 30, 2019, from https://techcrunch.com/2019/05/28/sf-will-allow-additional-dockless-bike-share-operators/

Divvy Bike Share. (2019). Divvy Data. Retrieved July 30, 2019, from https://www.divvybikes.com/system-data

Drynda, P. (2014). Development of bike share systems and their impact on the sustainability of urban transport . Case study of Opole Bike. *Central and Eastern European Journal of Management and Economics*, *2*(3), 199–215.

Faghih-Imani, A., Anowar, S., Miller, E. J., & Eluru, N. (2017). Hail a cab or ride a bike? A travel time comparison of taxi and bicycle-sharing systems in New York City. *Transportation Research Part A: Policy and Practice*, *101*, 11–21. https://doi.org/10.1016/j.tra.2017.05.006

Fishman, E. (2016). Bikeshare: A Review of Recent Literature. *Transport Reviews*, *36*(1), 92–113. https://doi.org/10.1080/01441647.2015.1033036

Fishman, E., Washington, S., & Haworth, N. (2014). Bike share's impact on car use: Evidence from the United States, Great Britain, and Australia. *Transportation Research Part D: Transport and Environment*, *31*, 13–20. https://doi.org/10.1016/j.trd.2014.05.013

Freund, D., Paul, A., Henderson, S. G., & Shmoys, D. B. (2017). *Data-Driven Rebalancing Methods for Bike-Share Systems*.

Fricker, C., & Gast, N. (2016). Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity. *EURO Journal on Transportation and Logistics*, *5*(3), 261–291. https://doi.org/10.1007/s13676-014-0053-5

Fuller, D., Gauvin, L., Kestens, Y., Daniel, M., Fournier, M., Morency, P., & Drouin, L. (2013). Impact evaluation of a public bicycle share program on cycling: A case example of BIXI in Montreal, Quebec. *American Journal of Public Health*, *103*(3), 85–92. https://doi.org/10.2105/AJPH.2012.300917

García-Palomares, J. C., Gutiérrez, J., & Latorre, M. (2012). Optimizing the location of stations in bike-sharing programs: A GIS approach. *Applied Geography*, *35*(1–2), 235–246. https://doi.org/10.1016/j.apgeog.2012.07.002

Goh, C. Y., Yan, C., & Jaillet, P. (2019). Estimating Primary Demand in Bike-sharing Systems. *SSRN Electronic Journal*, (February), 1–43. https://doi.org/10.2139/ssrn.3311371

Gold, M., Pinceti, S., & Federico, F. (2015). *2015 Environmental Report Card for Los Angeles County*. 99. Retrieved from www.environment.ucla.edu

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782. https://doi.org/10.1038/nature06958

Grenfell, B. T., Bjørnstad, O. N., & Kappey, J. (2001). Travelling waves and spatial hierarchies in measles epidemics. *Nature*, *414*(6865), 716–723. https://doi.org/10.1038/414716a

Gu, T., Kim, I., & Currie, G. (2019). To be or not to be dockless: Empirical analysis of dockless bikeshare development in China. *Transportation Research Part A: Policy and Practice*, *119*(August 2018), 122–147. https://doi.org/10.1016/j.tra.2018.11.007

Hahsler, M., & Hornik, K. (2007). TSP - Infrastructure for the traveling salesperson problem. *Journal of Statistical Software*, *23*(2), 1–21. https://doi.org/10.18637/jss.v023.i02

Hamilton, T. L., & Wichman, C. J. (2018). Bicycle infrastructure and traffic congestion: Evidence from DC's Capital Bikeshare. *Journal of Environmental Economics and Management*, *87*, 72–93. https://doi.org/10.1016/j.jeem.2017.03.007

Hinkley, D. V. (1971). Inference in Two-Phase Regression. *Journal of the American Statistical Association*, *66*(336), 736–743.

Hirsch, J. A., Stratton-Rayner, J., Winters, M., Stehlin, J., Hosford, K., & Mooney, S. J. (2019). Roadmap for free-floating bikeshare research and practice in North America. *Transport Reviews*, *39*(6), 706–732. https://doi.org/10.1080/01441647.2019.1649318

Horner, M. W., & O'Kelly, M. E. (2001). Embedding economies of scale concepts for hub network design. *Journal of Transport Geography*, *9*(4), 255–265. https://doi.org/10.1016/S0966-6923(01)00019-9

Huang, F., Qiao, S., Peng, J., & Guo, B. (2018). A Bimodal Gaussian Inhomogeneous Poisson Algorithm for Bike Number Prediction in a Bike-Sharing System. *IEEE Transactions on Intelligent Transportation Systems*, *PP*, 1–10. https://doi.org/10.1109/TITS.2018.2868483

Hufnagel, L., Brockmann, D., & Geisel, T. (2004). Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(42), 15124–15129. https://doi.org/10.1073/pnas.0308344101

Hulot, P., Aloise, D., & Jena, S. D. (2018). Towards Station-Level Demand Prediction for Effective Rebalancing in Bike-Sharing Systems. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 378–386. https://doi.org/10.1145/3219819.3219873

Hyland, M., Hong, Z., Karla, H., Farias, R. De, & Chen, Y. (2018). *Hybrid cluster-regression approach to model bikeshare station usage*. *115*(December 2017), 71–89.

Indego Bike Share. (2019). Indego Trip Data. Retrieved July 30, 2019, from https://www.rideindego.com/about/data

Institute for Transportation & Development Policy. (2018). The Bikeshare Planning Guide. *Institute for Transportation & Development Policy*.

Jäppinen, S., Toivonen, T., & Salonen, M. (2013). Modelling the potential effect of shared bicycles on public transport travel times in Greater Helsinki: An open data approach. *Applied Geography*, *43*, 13–24. https://doi.org/10.1016/j.apgeog.2013.05.010

Ji, Y., Ma, X., He, M., Jin, Y., & Yuan, Y. (2020). Comparison of usage regularity and its determinants between docked and dockless bike-sharing systems: A case study in Nanjing, China. *Journal of Cleaner Production*, *255*. https://doi.org/10.1016/j.jclepro.2020.120110

Jia, W., Tan, Y., Liu, L., Li, J., Zhang, H., & Zhao, K. (2019). Hierarchical prediction based on two-level Gaussian mixture model clustering for bike-sharing system. *Knowledge-Based Systems, 178*, 84–97. https://doi.org/10.1016/j.knosys.2019.04.020

Jiang, B., Yin, J., & Zhao, S. (2009). Characterizing the human mobility pattern in a large street network. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *80*(2), 1–11. https://doi.org/10.1103/PhysRevE.80.021136

Joh, K., Nguyen, M. T., & Boarnet, M. G. (2012). Can Built and Social Environmental Factors Encourage Walking among Individuals with Negative Walking Attitudes? *Journal of Planning Education and Research*, *32*(2), 219–236. https://doi.org/10.1177/0739456X11427914

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology and Evolution*, *19*(2), 101–108. https://doi.org/10.1016/j.tree.2003.10.013

Jurdak, R. (2013). The impact of cost and network topology on urban mobility: A study of public bicycle usage in 2 U.S. cities. *PLoS ONE*, *8*(11), 1–6. https://doi.org/10.1371/journal.pone.0079396

Kabra, A., Belavina, E., & Girotra, K. (n.d.). *Bike-Share Systems: Accessibility and Availability*. Retrieved from http://akabra.com/documents/Velib-shared.pdf

Keeling, M. J., Woolhouse, M. E., Shaw, D. J., Matthews, L., Chase-Topping, M., Haydon, D. T., … Grenfell, B. T. (2001). Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science*, Vol. 294, pp. 813–817. https://doi.org/10.1126/science.1065973

Kevin, E. P. & T. (2014). *2012 Seattle Community Greenhouse Gas Emissions Inventory*. (April). Retrieved from http://www.seattle.gov/Documents/Departments/OSE/2012 GHG inventory report_final.pdf

Kim, H. J., Fay, M. P., Feuer, E. J., & Midthune, D. N. (2000). Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in Medicine*, *19*(3), 335–351. https://doi.org/10.1002/(SICI)1097-0258(20000215)19:3<335::AID-SIM336>3.0.CO;2-Z

Kitamura, R., Chen, C., Pendyala, R. a M. M., & Narayanan, R. (2000). Micro-simulation of daily activity-travel patterns for travel. *Transportation*, *27*, 25–51. https://doi.org/10.1023/A:1005259324588

Kloimüllner, C., Papazek, P., Hu, B., & Raidl, G. R. (2014). Balancing Bicycle Sharing Systems: An Approach for the Dynamic Case. *European Conference on Evolutionary Computation in Combinatorial Optimization*, 73–84. https://doi.org/10.1007/978-3-662-44320-0_7

Koelbl, R., & Helbing, D. (2003). Energy and Scaling Laws in Human Travel Behaviour. *New Journal of Physics*, *5*(03), 1–48. https://doi.org/10.1088/1367-2630/5/1/348

Kou, Z., & Cai, H. (2019). Understanding bike sharing travel patterns: An analysis of trip data from eight cities. *Physica A: Statistical Mechanics and Its Applications*, *515*, 785–797. https://doi.org/10.1016/j.physa.2018.09.123

Lazarus, J., Pourquier, J. C., Feng, F., Hammel, H., & Shaheen, S. (2020). Micromobility evolution and expansion: Understanding how docked and dockless bikesharing models complement and compete – A case study of San Francisco. *Journal of Transport Geography*, *84*(May 2019). https://doi.org/10.1016/j.jtrangeo.2019.102620

LDA Consulting. (2016). *Capital Bikeshare 2016 Member Survey Report* (Vol. 20015).

Lee, K. L. K., Hong, S. H. S., Kim, S. J. K. S. J., Rhee, I. R. I., & Chong, S. C. S. (2009). SLAW: A New Mobility Model for Human Walks. *Ieee Infocom 2009*, 855–863. https://doi.org/10.1109/INFCOM.2009.5061995

Levine, R. V, & Norenzayan, A. (1999). The pace of life in 31 countries. *Journal of Cross-Cultural Psychology*, *30*(2).

Li, H., Zhang, Y., Ding, H., & Ren, G. (2019). Effects of dockless bike-sharing systems on the usage of the London Cycle Hire. *Transportation Research Part A: Policy and Practice*, *130*(December 2018), 398–411. https://doi.org/10.1016/j.tra.2019.09.050

Li, Y., Zheng, Y., Zhang, H., & Chen, L. (2015). Traffic prediction in a bike-sharing system. *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '15*, 1–10. https://doi.org/10.1145/2820783.2820837

Li, Y., Zheng, Y., Zhang, H., & Chen, L. (2016). *Traffic prediction in a bike-sharing system*. 1–10. https://doi.org/10.1145/2820783.2820837

Liang, X., Zheng, X., Lv, W., Zhu, T., & Xu, K. (2012). The scaling of human mobility by taxis is exponential. *Physica A: Statistical Mechanics and Its Applications*, *391*(5), 2135–2144. https://doi.org/10.1016/j.physa.2011.11.035

Lin, L., He, Z., & Peeta, S. (n.d.). *Predicting Station-level Hourly Demand in a Large-scale Bike-sharing Network: A Graph Convolutional Neural Network Approach*. https://doi.org/10.1016/j.trc.2018.10.011

Lin, L., He, Z., & Peeta, S. (2018). Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. *Transportation Research Part C: Emerging Technologies*, *97*(October), 258–276. https://doi.org/10.1016/j.trc.2018.10.011

Liu, J., Sun, L., Li, Q., Ming, J., Liu, Y., & Xiong, H. (2017). Functional Zone Based Hierarchical Demand Prediction For Bike System Expansion. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17*, 957–966. https://doi.org/10.1145/3097983.3098180

Liu, Z., Shen, Y., & Zhu, Y. (2018). Inferring Dockless Shared Bike Distribution in New Cities. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 378–386. https://doi.org/10.1145/3159652.3159708

Lobo, A. (2013). Bikes are faster door-to-door than cars or public transport within 5-10km. Retrieved November 18, 2017, from Better by Bicycle website: http://www.betterbybicycle.com/2013/12/bikes-are-faster-door-to-door-than-cars.html

Lobo, A. (2014). How accurate are Google Maps cycling time estimates? Retrieved January 1, 2018, from Better by Bicycle website: http://www.betterbybicycle.com/2014/09/how-accurate-are-google-maps-cycling.html

Luo, H., Kou, Z., Zhao, F., & Cai, H. (2019). Comparative life cycle assessment of station-based and dock-less bike sharing systems. *Resources, Conservation and Recycling*, *146*(March), 180–189. https://doi.org/10.1016/j.resconrec.2019.03.003

Luo, H., Zhao, F., Chen, W. Q., & Cai, H. (2020). Optimizing bike sharing systems from the life cycle greenhouse gas emissions perspective. *Transportation Research Part C: Emerging Technologies*, *117*(September 2019), 102705. https://doi.org/10.1016/j.trc.2020.102705

Malouff, D. (2017). All 119 US bikeshare systems, ranked by size. Retrieved November 16, 2017, from Greater Greater Washington website: https://ggwash.org/view/62137/all-119-us-bikeshare-systems-ranked-by-size

Mari, L., Bertuzzo, E., Righetto, L., Casagrandi, R., Gatto, M., Rodriguez-Iturbe, I., & Rinaldo, A. (2012). Modelling cholera epidemics: The role of waterways, human mobility and sanitation. *Journal of the Royal Society Interface*, *9*(67), 376–388. https://doi.org/10.1098/rsif.2011.0304

Martin, E. W., & Shaheen, S. A. (2014). Evaluating public transit modal shift dynamics in response to bikesharing: A tale of two U.S. cities. *Journal of Transport Geography*, *41*, 315–324. https://doi.org/10.1016/j.jtrangeo.2014.06.026

Mateo-Babiano, I., Bean, R., Corcoran, J., & Pojani, D. (2016). How does our natural and built environment affect the use of bicycle sharing? *Transportation Research Part A: Policy and Practice*, *94*, 295–307. https://doi.org/10.1016/j.tra.2016.09.015

McKenzie, G. (2019). Spatiotemporal comparative analysis of scooter-share and bike-share usage patterns in Washington, D.C. *Journal of Transport Geography*, *78*(May), 19–28. https://doi.org/10.1016/j.jtrangeo.2019.05.007

McMahon, J. (2019). Why Dockless Bike Share Doesn't Threaten Docked Bikes. Retrieved July 30, 2019, from https://www.forbes.com/sites/jeffmcmahon/2019/03/10/why-dockless-bike-share-doesnt-threaten-its-docked-ancestor/#60cdaf665c0c

Médard de Chardon, C., & Caruso, G. (2015). Estimating bike-share trips using station level data. *Transportation Research Part B: Methodological*, *78*, 260–279. https://doi.org/10.1016/j.trb.2015.05.003

Médard de Chardon, C., Caruso, G., & Thomas, I. (2017). Bicycle sharing system 'success' determinants. *Transportation Research Part A: Policy and Practice*, *100*, 202–214. https://doi.org/10.1016/j.tra.2017.04.020

Metro Bike Share. (2019). System Data. Retrieved October 27, 2019, from https://bikeshare.metro.net/about/data/

NACTO. (2017). *Bike Share in the U . S .: 2017*. Retrieved from https://nacto.org/bike-share-statistics-2017/

NACTO. (2020). *Shared MicroMobility in the US: 2019*. Retrieved from https://nacto.org/shared-micromobility-2019/

Nair, R., Miller-hooks, E., Hampshire, R. C., Bušić, A., Nair, R., Miller-hooks, E., … Bušić, A. (2013). *Large-Scale Vehicle Sharing Systems : Analysis of Vélib ' Large-Scale Vehicle Sharing Systems : Analysis*. *8318*. https://doi.org/10.1080/15568318.2012.660115

Nash, C. (2012). Genetics, Race, and Relatedness: Human Mobility and Human Diversity in the Genographic Project. *Annals of the Association of American Geographers*, *102*(3), 667–684. https://doi.org/10.1080/00045608.2011.603646

Newman, M. E. J. (2004). *Power laws, Pareto distributions and Zipf's law*. *7514*(September). https://doi.org/10.1016/j.cities.2012.03.001

O'Brien, O., Cheshire, J., & Batty, M. (2014). Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography*, *34*, 262–273. https://doi.org/10.1016/j.jtrangeo.2013.06.007

O'Mahony, E., & Shmoys, D. B. (2015). Data analysis and optimization for (Citi)bike sharing. *Proceedings of the National Conference on Artificial Intelligence*, *1*, 687–694.

Ome, C., & Latifa, O. (2014). Model-Based Count Series Clustering for Bike Sharing System Usage Mining : A Case Study with the Vélib System of Paris. *ACM Transactions on Intelligent Systems and Technology*, *5*(3), 1–21. https://doi.org/10.1145/2560188

Pan, L., Cai, Q., Fang, Z., Tang, P., & Huang, L. (2019). A Deep Reinforcement Learning Framework for Rebalancing Dockless Bike Sharing Systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*. Retrieved from http://arxiv.org/abs/1802.04592

Park, C., & Sohn, S. Y. (2017). An optimization approach for the placement of bicycle-sharing stations to reduce short car trips : An application to the city of Seoul. *Transportation Research Part A*, *105*(June), 154–166. https://doi.org/10.1016/j.tra.2017.08.019

Pasion, C., Oyenuga, C., Gouin, K., & LLC, C. (2017). *Inventory of New York City greenhouse gas emissions in 2015*. (April), 1–56.

Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic Biology*, *53*(5), 793–808. https://doi.org/10.1080/10635150490522304

Program, C. (2015). *2015 San Francisco Geographic Greenhouse Gas*. (May).

Ramos-Fernández, G., Mateos, J. L., Miramontes, O., Cocho, G., Larralde, H., & Ayala-Orozco, B. (2004). Lévy walk patterns in the foraging movements of spider monkeys (Ateles geoffroyi). *Behavioral Ecology and Sociobiology*, *55*(3), 223–230. https://doi.org/10.1007/s00265-003-0700-6

Ratcliffe, J. H. (2012). The Spatial Extent of Criminogenic Places: A Changepoint Regression of Violence around Bars. *Geographical Analysis*, *44*(4), 302–320. https://doi.org/10.1111/j.1538-4632.2012.00856.x

Report, F. (2017). *CITY OF CHICAGO GREENHOUSE GAS INVENTORY REPORT CALENDAR YEAR 2015*. Chicago. Retrieved from https://www.cityofchicago.org/content/dam/city/progs/env/GHG_Inventory/CityofChicago_2015_GHG_Emissions_Inventory_Report.pdf

Rhee, I., Rhee, I., Shin, M., Shin, M., Hong, S., Lee, K., … Chong, S. (2011). On the Levy-walk Nature of Human Mobility: Do Humans Walk like Monkeys? *IEEE/ACM Transactions on Networking*, *19*(3), 630–643. https://doi.org/10.1109/TNET.2011.2120618

Rinde, M. (2019). Dockless e-bike share coming to Philly in 2019. Retrieved July 30, 2019, from http://planphilly.com/articles/2019/04/17/dockless-e-bike-share-coming-to-philly-in-2019

Romero, J. P., Ibeas, A., Moura, J. L., Benavente, J., & Alonso, B. (2012). *A simulation-optimization approach to design efficient systems of*. *54*, 646–655. https://doi.org/10.1016/j.sbspro.2012.09.782

Rosenkrantz, D. J., Stearns, R. E., & Lewis, P. M. (2009). An analysis of several heuristics for the traveling salesman problem. *Fundamental Problems in Computing: Essays in Honor of Professor Daniel J. Rosenkrantz*, *6*(3), 45–69. https://doi.org/10.1007/978-1-4020-9688-4_3

Schuijbroek, J., Hampshire, R. C., & van Hoeve, W. J. (2017). Inventory rebalancing and vehicle routing in bike sharing systems. *European Journal of Operational Research*, *257*(3), 992–1004. https://doi.org/10.1016/j.ejor.2016.08.029

Shaheen, S. a., Cohen, A. P., & Martin, E. W. (2013a). Public Bikesharing in North America: Early Operator Understanding and Emerging Trends. *Transportation Research Record: Journal of the Transportation Research Board*, *2387*, 83–92. https://doi.org/10.3141/2387-10

Shaheen, S., Guzman, S., & Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia. *Transportation Research Record: Journal of the Transportation Research Board*, *2143*, 159–167. https://doi.org/10.3141/2143-20

Shaheen, S., Martin, E., & Cohen, A. (2013b). Public Bikesharing and Modal Shift Behavior: A Comparative Study of Early Bikesharing Systems in North America. *International Journal of Transportation*, *1*(1), 35–54. https://doi.org/10.14257/ijt.2013.1.1.03

Singhvi, D., Singhvi, S., Frazier, P. I., Henderson, S. G., Mahony, E. O., Shmoys, D. B., & Woodard, D. B. (2015). Predicting Bike Usage for New York City ' s Bike Sharing System. *File:///Users/Spc/ARCHIVO/BIBLIOTECA/SHARING BIKES/1-S2.0-S187704281400086X-Main.Pdf*, 110–114.

Singla, A., Santoni, M., Bartok, G., Mukerji, P., Meenen, M., & Krause, A. (2015). Incentivizing Users for Balancing Bike Sharing Systems. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence Pattern*, (June 2014), 723–729.

Sisson, P. (2017). New bike-share system promises 'dockless without the drawbacks.' Retrieved July 30, 2019, from https://www.curbed.com/2017/11/30/16720066/bike-share-dockless-pace-cities-cycling

Song, C., Koren, T., Wang, P., & Barabási, A.-L. (2010). Modelling the scaling properties of human mobility. *Nature Physics*, *6*(10), 818–823. https://doi.org/10.1038/nphys1760

Spring, P. (2015). *PHILADELPHIA CITYWIDE GREENHOUSE GAS INVENTORY , 2012 Table of Contents*.

Susan, S., Elliot, M., Nelson, C., Adam, C., & Mike, P. (2014). *Public Bikesharing in North America During a Period of Rapid Expansion : Understanding Business Models , Industry Trends and User Impacts*.

Syarif, I., Prugel-Bennett, A., & Wills, G. (2016). SVM parameter optimization using grid search and genetic algorithm to improve classification performance. *Telkomnika (Telecommunication Computing Electronics and Control)*, *14*(4), 1502–1509. https://doi.org/10.12928/TELKOMNIKA.v14i4.3956

Tao, S., & Pender, J. (2017). A Stochastic Analysis of Bike Sharing Systems. *ArXiv Preprint*, (arXiv:1708.08052). Retrieved from http://arxiv.org/abs/1708.08052

The City of Chicago. (2019). Dockless Bike Share Pilot Project. Retrieved September 25, 2019, from https://www.chicago.gov/city/en/depts/cdot/supp_info/dockless-bike-share-pilot-project.html

Thompson, J., Potok, N., & Blumerman, L. (2016). *TIGER / Line ® Shapefiles Technical Documentation*.

U.S.DOT. (2017). 2017 National Household Travel Survey. Retrieved from http://nhts.ornl.gov

United Nations Environment Programme (UNEP). (2018). *The Weight of Cities: Resource requirements of future urbanization*. Retrieved from www.internationalresourcepanel.org

USDOT. (2009). *Summary of travel trend: 2009 National Household Travel Survey*. Retrieved from http://nhts.ornl.gov

Veloso, M., & Phithakkitnukoon, S. (2011). Exploratory study of urban flow using taxi traces. *First Workshop on …*, (December 2015), 1–8. Retrieved from http://www.researchgate.net/publication/232175450_Exploratory_Study_of_Urban_Flow_using_Taxi_Traces/file/79e41507829c808c03.pdf

Viswanathan, G. M., Afanasyev, V., Buldyrev, S. V., Murphy, E. J., Prince, P. A., & Stanley, H. E. (1996). Lévy flight search patterns of wandering albatrosses. *Nature*, Vol. 381, pp. 413–415. https://doi.org/10.1038/381413a0

Wang, B., & Kim, I. (2018). Short-term prediction for bike-sharing service using machine learning. *Transportation Research Procedia*, *34*, 171–178. https://doi.org/10.1016/j.trpro.2018.11.029

Wang, J., & Lindsey, G. (2019). Do new bike share stations increase member use: A quasi-experimental study. *Transportation Research Part A: Policy and Practice*, *121*(December 2017), 1–11. https://doi.org/10.1016/j.tra.2019.01.004

Wang, J., Xu, J., Zhao, C., Peng, Y., & Wang, H. (2019). An ensemble feature selection method for high-dimensional data based on sort aggregation. *Systems Science and Control Engineering*, *7*(2), 32–39. https://doi.org/10.1080/21642583.2019.1620658

Watson, B. C., & Telenko, C. (2019). Predicting Demand of Distributed Product Service Systems by Binomial Parameter Mapping: A Case Study of Bike Sharing Station Expansion. *Journal of Mechanical Design, Transactions of the ASME*, *141*(10). https://doi.org/10.1115/1.4043366

Xu, C., Ji, J., & Liu, P. (2018a). The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. *Transportation Research Part C*, *95*(July), 47–60. https://doi.org/10.1016/j.trc.2018.07.013

Xu, X., Ye, Z., Li, J., & Xu, M. (2018b). Understanding the Usage Patterns of Bicycle-Sharing Systems to Predict Users' Demand: A Case Study in Wenzhou, China. *Computational Intelligence and Neuroscience*, *2018*, 1–21. https://doi.org/10.1155/2018/9892134

Yan, X.-Y., Han, X.-P., Wang, B.-H., & Zhou, T. (2013). Diversity of individual mobility patterns and emergence of aggregated scaling laws. *Scientific Reports*, *3*, 2678. https://doi.org/10.1038/srep02678

Yang, Z., Chen, J., Hu, J., Shu, Y., & Cheng, P. (2019). Mobility Modeling and Data-Driven Closed-Loop Prediction in Bike-Sharing Systems. *IEEE Transactions on Intelligent Transportation Systems*, 1–12. https://doi.org/10.1109/tits.2018.2886456

Yang, Z., Hu, J., Shu, Y., Cheng, P., Chen, J., & Moscibroda, T. (2016). Mobility Modeling and Prediction in Bike-Sharing Systems. *MobiSys*, 165–178.

Zhang, J., Pan, X., Li, M., & Yu, P. S. (2016a). Bicycle-sharing systems expansion: Station re-deployment through crowd planning. *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*. https://doi.org/10.1145/2996913.2996926

Zhang, J., Zheng, Y., Qi, D., Li, R., & Yi, X. (2016b). *DNN-Based Prediction Model for Spatio-Temporal Data*.

Zhang, Y., Lin, D., & Mi, Z. (2019). Electric fence planning for dockless bike-sharing services. *Journal of Cleaner Production*, *206*, 383–393. https://doi.org/10.1016/j.jclepro.2018.09.215

Zhang, Y., & Mi, Z. (2018). Environmental benefits of bike sharing: A big data-based analysis. *Applied Energy*, *220*(June), 296–301. https://doi.org/10.1016/j.apenergy.2018.03.101

Zhang, Y., Thomas, T., Brussel, M. J. G., & Van Maarseveen, M. F. A. M. (2016c). Expanding bicycle-sharing systems: Lessons learnt from an analysis of usage. *PLoS ONE*, *11*(12), 1–25. https://doi.org/10.1371/journal.pone.0168604

Zhou, X. (2015). Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in Chicago. *PLoS ONE*, *10*(10), 1–20. https://doi.org/10.1371/journal.pone.0137922

Zhou, Y., Chen, H., Li, J., Wu, Y., Wu, J., & Chen, L. (2019). Large-Scale Station-Level Crowd Flow Forecast with ST-Unet. *ISPRS International Journal of Geo-Information*, *8*(3), 140. https://doi.org/10.3390/ijgi8030140

VITA

Zhaoyu Kou

EDUCATION

Purdue University, West Lafayette, IN
Ph.D. in Industrial Engineering, GPA: 4.0/4.0 *December 2020*
Advisor: Dr. Hua Cai (2016 – present)
Dissertation title: "*Bike Share System Modeling: Travel Patterns, Environmental Benefits, System Expansion, and Impacts of System Types"*

Columbia University, New York, NY
M.S. in Earth Resources Engineering *Feb 2016*
Advisor: Dr. Vasilis M. Fthenakis (2014 – 2016)
Thesis title: *"Eco-design of e-waste recycling networks by multi-objective optimization"*

Tsinghua University, Beijing, China
B.S. in Environmental Engineering *July 2014*
Advisor: Zongguo Wen (2013-2014)
Thesis title: *"Development of a municipal-renewable-resources-collection system based on Internet of Things"*

AWARDS AND HONORS

- Hugh W. and Edna M. Donnan Dissertation Fellowship, 2019-2020
- Honorable Mention, Student Poster Competition, Sustainable Urban Systems Section, 2019 International Society for Industrial Ecology (ISIE) Biannual Conference, July 9, 2019, Beijing, China.
- College of Engineering and Graduate School Scholarship, Purdue University, 2016 & 2018
- Veolia Scholarship, Tsinghua University, 2011

PROFESSIONAL EXPERIENCE

Purdue University, West Lafayette, IN *August 2016 - Present*
Research Assistant

Viral Launch, Indianapolis, IN *May 2019 – August 2019*
Data Scientist Intern

Purdue University, West Lafayette, IN *August 2016 – December 2018*
Teaching Assistant of course IE343 - "Engineering Economics"

Everbright Prestige Capital, Beijing, China *March 2016 – July 2016*
Investment Group Intern

Clowder, New York, NY *May 2015 - December 2015*
Operations and Supply Chain Management Intern

Friends of Nature (FON), Beijing, China *July 2013 - August 2013*
Public Affairs Project Assistant

The Energy Foundation, Beijing, China *January 2013 - February 2013*
Environmental Management Program Intern

PROFESSIONAL AFFILIATION & SERVICE

Reviewer (Journal Manuscripts)

- Journal of Cleaner Production
- Physica A: Statistical Mechanics and its Applications
- Resources, Conservation and Recycling

Volunteer

- National Shared Mobility Summit, 2018

Membership

- Institute of Industrial and Systems Engineers (IISE), Student Member, 2018 – present
- Institute for Operations Research and the Management Sciences (INFORMS), Student Member, 2019 – present

STUDENT MENTORING

Mentored the following students for their research projects at Purdue University

Master Students

- Zhuoli, Yin, Industrial Engineering
- Ruihua Sun, Industrial Engineering
- Hao Luo, Environmental and Ecological Engineering

Undergraduate Students

- Weizhou Zhang, Industrial Engineering
- Mokammel Hossain Sanju, Environmental and Ecological Engineering
- Ana M Diaz Abad, Industrial Engineering
- Xiang Feng, Industrial Engineering

- Delzin Gamir, Industrial Engineering
- Anna Poznyak, Industrial Engineering
- Fajar Ausri, Industrial Engineering
- Jiaqi Zhu, Computer Science

GRANTS

- Symposium on Industrial Ecology for Young Professionals (SIEYP) Travel Scholarship, 2019 ($1,000)
- Purdue Climate Change Research Center Travel Grant, 2018 ($800)
- Gordon Research Conference (GRC) Travel Grant, 2018 ($1,250)
- Gordon Research Seminar (GRS) Travel Grant, 2018 ($295)
- ISIE Student Chapter Travel Scholarship, 2017 ($670)
- Purdue Climate Change Research Center Travel Grant, 2017 ($500)

PUBLICATIONS AND SCHOLARLY WORK

Published Papers

- **Kou, Z.**, & Cai, H. (2019). Understanding bike sharing travel patterns: An analysis of trip data from eight cities. *Physica A: Statistical Mechanics and its Applications*, 515, 785-797.
- Luo, H., **Kou, Z.**, Zhao, F., & Cai, H. (2019). Comparative life cycle assessment of station-based and dock-less bike sharing systems. *Resources, Conservation and Recycling*, 146, 180-189.
- **Kou, Z.**, Wang, X., Chiu, S. F. A., & Cai, H. (2020). Quantifying greenhouse gas emissions reduction from bike share systems: a model considering real-world trips and transportation mode choice patterns. *Resources, Conservation and Recycling*, 153, 104534.

Works under Review and in Progress

- **Kou, Z.**, & Cai, H. (TBD). Comparing the performance of different types of bike share systems. *Transportation Research Part D: Transport and Environment* (under review).
- **Kou, Z.**, & Cai, H. (TBD). The station interactions in bike share system expansion and their application to the demand prediction. *Transportation planning and technology* (under preparation).
- Sun, R., **Kou, Z.**, & Cai, H. (TBD). Estimating the rebalancing demands and vehicle routing for existing bike share systems. *Transportation Research Part C: Emerging Technologies* (under preparation).
- Yin, Z., **Kou, Z.**, & Cai, H. (TBD). Dynamic Bike Sharing Reposition by Deep Reinforcement Learning. *Journal of Machine Learning Research* (under preparation).

Presentations at Conferences

- **Kou, Z.**, & Cai, H., "Demand Prediction For Bike Share System Expansion and a Modeling Framework to Evaluate System Performance: a Case Study of Chicago", Platform presentation, INFORMS 2019 annual meeting, Oct. 22, Seattle WA
- **Kou, Z.**, & Cai, H., ″Comparing dock-based, dockless, and hybrid bike share systems regarding bike supply: a case study of Chicago″, Poster presentation, 2019 ISIE Conference, July 9, 2019, Beijing, China. (Received **Honorable Mention** of the ISIE 2019 Student Poster Competition from the Sustainable Urban Systems section.)
- **Kou, Z.**, & Cai, H., "How many bikes are needed in a bike share system?", Poster presentation, 2019 Association of Environmental Engineering & Science Professors (AEESP) Research and Education Conference, Tempe, AZ, May 14- 16, 2019
- **Kou, Z.**, & Cai, H., "Quantifying the Environmental Benefits of Bike Share Systems", Poster presentation, 2019 Transportation Research Board (TRB) Annual Conference, January 14, 2019, Washington, D.C.
- **Kou, Z.**, & Cai, H., "Understanding the Sustainability of Bike Sharing Systems: A Tale of Eight Cities", Poster presentation, Industrial Ecology Gordon Research Conference, May 22, 2018, Les Diablerets, Switzerland
- **Kou, Z.**, & Cai, H., "Understanding the Sustainability of Bike Sharing Systems: A Tale of Eight Cities" Platform presentation, ISIE- ISSST Joint Conference, June 29, 2017, Chicago, IL
- **Kou, Z.**, & Cai, H., "Understanding the Sustainability of Bike Sharing Systems: A Tale of Eight Cities", Poster presentation, 2017 Association of Environmental Engineering & Science Professors (AEESP) Research and Education Conference, June 21, 2017, Ann Arbor, MI 47907