FINAL REPORT

Project F2

May 2021

Discovering Potential Market for the Integration of Public Transportation & Emerging Shared-Mobility Services

Dr. Lili Du | University of Florida

Dr. Xia Jin | Florida International University

Dr. William J. Davis, P.E. | The Citadel

Jiahua Qiu | University of Florida

Wang Peng | University of Florida

Ghazaleh Azimi | Florida International University

Alireza Rahimi | Florida International University

Dr. Ming Lee | Florida International University

STRIDE | Southeastern Transportation Research, Innovation, Development and Education Center

UF | Transportation Institute UNIVERSITY of FLORIDA

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. Project F2	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Discovering Potential Market for the Integration of Public Transportation and Emerging Shared-Mobility Services		**5. Report Date** May 7, 2021 **6. Performing Organization Code**
7. Author(s) Dr. Lili Du, University of Florida Dr. Xia Jin, Florida International University Jiahua Qiu (graduate student), University of Florida Peng Wang (graduate student), University of Florida Ghazaleh Azimi (graduate student), Florida International University Alireza Rahimi (graduate student), Florida International University Dr. Ming Lee, Florida International University		**8. Performing Organization Report No.** Project F2
9. Performing Organization Name and Address University of Florida/Civil & Coastal Engineering 365 Weil Hall, PO Box 116580 Gainesville, FL 32611 Florida International University/Civil & Environmental Engineering 10555 W. Flagler Street, EC 3680 Miami, FL 33174		**10. Work Unit No.** **11. Contract or Grant No.** Funding Agreement Number 69A355174710
12. Sponsoring Agency Name and Address University of Florida Transportation Institute Southeastern Transportation Research, Innovation, Development and Education Center 365 Weil Hall, P.O. Box, Gainesville, FL 32611 U.S Department of Transportation/Office of Research, Development & Tech 1200 New Jersey Avenue, SE Washington, DC 20590 United States		**13. Type of Report and Period Covered** 1/15/2019 to 5/7/2021 **14. Sponsoring Agency Code**

15. Supplementary Notes

16. Abstract

This project conducted a comprehensive data analysis to answer two questions: 1) Who are the potential demands with high probability to use intermodal services provided by hybrid systems? (Task 1); 2) Where and when are the supply gaps to coordinate public transit services with shared-mobility service? (Task 2). To accomplish these research objectives, the project team draws from several data sources and approaches to conduct research from both the demand and supply perspectives. Specifically, Task 1 of this study investigated the influential factors that affect transit users' choices of access and egress modes, including TNC or taxi, drive alone (PNR), carpool (KNR, carpool or shuttle), and micromobility modes (bike-sharing, scooters, etc.), using a transit on-board survey conducted in Spring 2017 for the Orlando metropolitan area. Task 2 of this study used the transit trip data and ridesharing trajectory data in the second ring region of Chengdu, China to develop innovative data analysis and machine learning approaches to explore the transit service gaps in both flexibility and coverage. Together, these efforts provide a snapshot to better understand the potential service gaps and demands (such as first/last mile gaps and demand) for promoting hybrid mobility that integrates shared mobility and public transit. Overall, the outputs of this project will increase the use of sustainable transportation modes, which may reduce urban congestion, emission, and energy consumptions. Thus, the success of this project will help establish an eco-friendly transportation system.

17. Key Words	18. Distribution Statement
Ridesharing-transit hybrid urban public mobility, supply-demand market, first/last mile, trip data, access/egress mode	No restrictions

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
N/A	N/A	83 Pages	N/A

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

ACKNOWLEDGEMENT OF SPONSORSHIP AND STAKEHOLDERS

This work was sponsored by a contract from the Southeastern Transportation Research, Innovation, Development and Education Center (STRIDE), a Regional University Transportation Center sponsored by a grant from the U.S. Department of Transportation's University Transportation Centers Program

Funding Agreement Number - 69A3551747104

LIST OF AUTHORS

Lead PI:

Lili Du, Ph.D.
University of Florida
Email address: lilidu@ufl.edu
ORCID Number: 0000-0003-1740-1209

Co-PI:

Xia Jin, Ph.D., AICP
Florida International University
Email address: xjin1@fiu.edu
ORCID Number: 0000-0002-8660-3528

Additional Researchers:

Jiahua Qiu
University of Florida
Email address: jq22@ufl.edu
ORCID Number: 0000-0002-3025-5878

Peng Wang
University of Florida
Email address: pengw@ufl.edu
ORCID Number: 0000-0002-0269-7310

Ghazaleh Azimi
Florida International University
Email address: gazimi@fiu.edu
ORCID Number: 0000-0001-5646-6908

Alireza Rahimi
Florida International University
Email address: alrahimi@fiu.edu
ORCID Number: 0000-0002-4555-7799

Ming Lee, MD, Ph.D., PE
Florida International University
Email address: milee@fiu.edu
ORCID Number: 0000-0001-6856-6367

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The advancement of communication and information technology has enabled travelers to request, track, and pay for trips via mobile devices. This significant convenience promotes emerging travel modes such as shared-mobility service including carsharing, bikesharing, ridesharing (like Uber and Lyft), and private shuttles (like Bay-Area tech shuttles). Shared-mobility owners have claimed that these new traffic modes will help reduce car ownership and promote the ridership of public transit, while transit agencies, often unsure of how to coexist with them, express concerns on the potential competition and potential extra traffic and associated congestion. These mixed opinions raise the urgent need to fully understand the potential supply-demand market so that we foster cooperation between public transit and shared mobility, taking the advantage of both. Motivated by this view, this project conducted a comprehensive data analysis to answer two questions: 1) Who are the potential demands with high probability to use intermodal services provided by hybrid systems? (Task 1); 2) Where and when are the supply gaps to coordinate public transit services with shared-mobility service? (Task 2). To accomplish these research objectives, the project team draws from several data sources and approaches to conduct research from both the demand and supply perspectives. Specifically, Task 1 of this study investigated the influential factors that affect transit users' choices of access and egress modes, including TNC or taxi, drive alone (PNR), carpool (KNR, carpool or shuttle), and micromobility modes (bike-sharing, scooters, etc.), using a transit on-board survey conducted in Spring 2017 for the Orlando metropolitan area. Task 2 of this study used the transit trip data and ridesharing trajectory data in the second ring region of Chengdu, China to develop innovative data analysis and machine learning approaches to explore the transit service gaps in both flexibility and coverage. Together, these efforts provide a snapshot to better understand the potential service gaps and demands (such as first/last mile gaps and demand) for promoting hybrid mobility that integrates shared mobility and public transit. Overall, the outputs of this project will increase the use of sustainable transportation modes, which may reduce urban congestion, emission, and energy consumptions. Thus, the success of this project will help establish an eco-friendly transportation system.

Keywords:
Ridesharing-transit hybrid urban public mobility, supply-demand market, first/last mile, trip data, access/egress mode

EXECUTIVE SUMMARY

Emerging shared mobility services such as Uber and Lyft have quickly spread in popularity and may have both positive and negative impacts on public transit. Hybrid public mobility service systems that incorporate both modes offer a promising win-win solution. However, we still lack knowledge of when and where to integrate them and who needs such services. Namely, the spatiotemporal supply-demand market is unclear. To make up this gap, the project team draws from several data sources and approaches to conduct research from both demand and supply perspectives. The main findings are summarized as follows. One common interest of both demand and supply side studies is the first/last mile (FLM) gap of transit service.

Task 1 studied transit on-board survey data for the Orlando metropolitan area and used a Smart Location Database (SLD, comprehensive land-use attributes) to investigate how land use characteristics may contribute to users' choice for access and egress modes, beyond the personal and household attributes. Employing separate multinomial logit models, the study denotes the following key findings: (i) Trips showing higher potential demand of TNCs for FLM purposes include airport or university trips, trips with longer access distance, trips made by persons with higher household income; (ii) Trips with less potential of using TNC for FLM purposes include sports events, medical visits, visitor trips, and evening trips; (iii) In terms of the impacts of land-use attributes, higher employment and household entropy and higher diversity at the origin showed positive impacts on the use of micro-mobility and walking, and reduced the probability of using motorized modes, including TNCs, for first mile purposes. On the destination side, higher diversity seemed to encourage the use of TNCs and drive alone modes for last-mile purposes.

Task 2 studied transit and ridesharing trip data (including O-D information) collected in the second ring region of Chengdu, China. We considered ridesharing trip data as the detectors to demonstrate the deficiency of the flexibility and coverage in the existing transit system. By mashing the 3D space spinning by the trip data to an optimal 3D discretization grid, and measuring the heat in each cube by the bus or ridesharing service rate, we conducted a heatmap analysis and obtained four important findings: (i) The ridesharing service swarm (RS) areas are the potential locations to implement new micro-transit services or transit lines and stations; (ii) The areas consisted of two ridesharing service swarm zones sandwiched by a transit service zone (“sandwich” pattern) revealing the potential first/last mile (FLM) service gaps, where response-on-demand services are needed; (iii) Land use data indicate that RS and FLM zones are often commercial centers and large residential areas; and (v) Both the RS zones and FLM zones evolved over time. Our deep learning method is effectively able to predict the variation.

This study provides useful insights into the deficiency of service coverage/flexibility in existing transit systems and the factors that may influence transit users’ choice of modes for access and egress purposes. It may help transit agencies and planners in understanding the potential supply and demand market for refining existing transit systems considering the integration of response-on-demand mobility services. The findings derived from Tasks 1 and 2 may be limited

to the dataset collected from different cities, the second ring area of Chengdu and the Orlando metropolitan area. We propose future research for conducting a comprehensive case study based on a complete dataset in a city and investigating the transferability of the findings.

1.0 TASK 1: ANALYSIS FOR DEMAND MARKET

1.1 INTRODUCTION

Rapidly evolving advanced technologies such as autonomous and connected vehicles, together with shared mobility services (such as carsharing or ridesourcing), may forever change how people live and travel. At the same time, major demographic and societal trends across America may transform future urban mobility. Baby Boomers and Millennials (also called the Gen Ys), the two largest generations in the U.S. – at 76 million and 80 million respectively, could bring huge opportunities for public transportation over the next two decades.

Recent surveys have shown an increasing use of new mobility services for first/last mile connection to transit. However, limited research has been conducted to quantify their impacts on the existing public transportation service and to identify empirical methods to forecast these impacts on the future integration of new shared mobility services with transit services. There is a pressing need to understand how emerging mobility options may reshape the way people travel and how public transportation may find new opportunities to serve their mobility needs.

This project aims to assess the potential of integrating transit and shared mobility services and examine how emerging mobility options and vehicle technologies may work together to influence the transit market. Although many have discussed the future of transit in light of the emerging technologies and trends, very little is known on how to capture and quantify the impacts on the transit market. The results of this study will provide important inputs for public transit agencies and private service providers to formulate regulations and policies and develop business models that enable the creation of integrated, multimodal, and sustainable mobility systems that embrace the emerging technologies and advancements.

1.1.1 Objective

This project assesses the potential impacts of emerging technologies and mobility services on transit market share. The objectives are to

1. Investigate existing market of shared mobility as access and egress mode for transit, including rider characteristics (income, age group, education, and car ownership, etc.), travel patterns (origin/destination type, trip length, access/egress distance, etc.), and other modal features (trip purpose, rail/bus preferences, etc.) and to
2. Explore the spatial pattern of transit trips by different access/egress mode, and evaluate the impacts of land use and built environment factors on the use of shared mobility for first/last-mile connections.

1.1.2 Scope

This project focuses on the most recent transit on-board surveys conducted in the Orlando metropolitan area in 2017. Various land-use data are also integrated with the survey data for

spatial analysis. These data include the EPA Smart Location Database, the U.S. Census TIGER All Roads data, and the General Transit Feed Specification (GTFS) data.

1.2 LITERATURE REVIEW

Various researchers have studied the potential effects of shared mobility services and autonomous vehicle (AV) technologies on public transit ridership. This section focuses on those that specifically looked at the potential of integrating these options with transit service for first/last mile solutions.

Jaller et al. (2019) evaluated the benefits of a first-mile transit access program using shared mobility services. The potential demand shifts from drive-alone mode to the proposed program were investigated. A simulation and optimization framework was developed and implemented in the San Francisco Bay Area for access to Bay Area Rapid Transit (BART). Results showed that by assuming a 25% reduction in travel time, about 18% of increased AM work trips moved from drive-alone mode to the simulated mode. Moreover, total vehicle miles traveled (VMT) decreased dramatically in simulated mode, while the generalized cost for trips increased significantly due to significant travel time increases.

Vakayil et al. (2017) integrated an autonomous mobility-on-demand (AMoD) service with mass transit service so that AMoD functions as a first-and-last-mile solution. They applied the simulated system on Washington DC using car2go user trips and hub/frequency data from the DC Metro system. The results for comparison of AMoD and integrated system (AMoD-Transit) revealed that most of the trips consist of transit segments during rush hours, but only a very small percentage of trips are served completely via transit. Moreover, the integrated system provides a 50% reduction in total VMT, improved mobility, and decreased the number of walkways, especially during rush hours.

Farhan et al. (2018) evaluated the operation of Shared Autonomous Electric Vehicles (SAEVs) for various vehicle range and charging infrastructure by using a simulation model for the first/last mile problem. They applied the proposed simulation model on Tukwila Station operations in the Seattle metropolitan area. The data included the 2016 origin/destination survey of light-rail riders. Results demonstrated that the SAEV fleet could be used as a first/last mile solution increasing mobility and decreasing total VMT through ridesharing. Moreover, utilizing fast charging technology and long-range vehicles effectively reduces the SAEV fleet and wait time.

In a study by Shen et al. (2017), the potential effects of on-demand sharing autonomous vehicle (SAV) on public transit were investigated. They applied an on-demand AV sharing service instead of low-demand public transit to analyze its effects on the first/last mile problem focusing on workdays. The proposed framework was applied in Tampines Town in Singapore. Also, an agent-based simulation model was utilized to evaluate the performance of the integrated service. Results showed that by sharing the last-mile rides and careful AV fleet

selection, the integrated service could reduce the average passengers' out-of-vehicle time, reduce the occupancy of road resources (VMT), and increase the possibility of financial viability.

Berrada et al. (2019) investigated the potential demand for autonomous taxis (aTaxis) while considering interactions with transit. They integrated a dynamic supply model (aTaxis) into a static demand model (scheduled services) and applied the proposed framework to Palaiseau, a French city located in the Paris metropolitan area. Results demonstrated that aTaxis were more attractive for relatively short trips (average trip length of 4 km). The study indicated that the introduction of aTaxies improved service quality and reduced the usage of private cars. For markets not directly served by bus rapid transit (BRT), aTaxis also significantly reduced the cost of users.

Alemi and Rodier (2018) investigated the potential market demand for a first-mile transit access program in the San Francisco Bay Area through agent-based demand and supply modeling. The study focused on drive alone commuters, who had the opportunity to take the rail to work. Results showed that by switching to a TNC taxi and taking BART to work, 31% of drive-alone trips could reduce their generalized travel cost. Moreover, if all commuters shifted to TNC and BART, total VMT could be decreased by 0.5 million miles during the morning commute period. Taking into consideration the cost saved by the integrated network, the new service would be beneficial for low-income households with few cars.

In a study by Pinto et al. (2018), the impacts of first-mile SAV service on transit demand in the suburban area were investigated. Multinomial logit model and dynamic traffic assignment models were integrated and implemented in the region served by Chicago Transit Authority (CTA). Agent-based micro-simulation tools for modeling the movements of travelers and SAVs were developed. Results indicated that the integration of SAV to the current transit network reduced the number of driving trips. In different scenarios, the study observed both substitution effects (some of the transit trips shifted to SAV) and complement effects (transit became more desirable when integrated with SAV).

Davidson et al. (2017) investigated the pattern of Uber requests and the implications for transit use. They compared Uber trip origins from two datasets, publicly released Uber origin data through the New York City Taxi and Limousine Commission, and Uber requests summoned through a Transit app. Results showed that a higher frequency of Uber requests was recorded from the transit app users than the general requests within 250 feet of a transit station. Uber requests through the transit apps also showed greater dispersion throughout the city. This indicated that Uber served as a viable option to make up gaps in existing transit options.

Stiglic et al. (2018) investigated the potential benefits of ridesharing and public transit integration, as well as required ride-matching technologies to support this system. They utilized ride-matching algorithms to find feasible matches. The proposed framework was implemented on a rectangular metropolitan area of 20 by 10 miles covering a circular urban center with a radius of 2.5 miles and a sprawling suburban area. The results indicated that the proposed system significantly improved mobility and increased the use of public transit. Moreover, total

system wide VMT were reduced, and so were the negative externalities associated with car travel.

Yan et al. (2018) evaluated traveler responses to a proposed integrated transit system and ridesourcing services named MTransit, at the University of Michigan Ann Arbor campus. A survey collecting both revealed preference (RP) and stated preference (SP) data was conducted, and a RS-SP mixed logit model was fitted. The mixed logit model outputs were applied to predict demand for MTransit under different scenarios. Results indicated that passengers were discouraged from using MTransit mainly because of transfers and additional pickups. Moreover, it was found that if low-ridership bus lines were replaced with ridesourcing services, transit ridership could increase slightly, and operational costs could decrease. Finally, they found that when used to provide convenient last-mile connections, ridesourcing could provide a significant boost to transit.

Hall et al. (2018) investigated whether Uber had substitute or complement effects on ridership. They applied a difference-in-differences approach on a dataset including transit ridership data, Uber entry and exit, and a variety of controls for 2004–2015. The results indicated that Uber's effect varied based on the Metropolitan Statistical Areas (MSA) population, transit ridership, and Uber penetration. It was found that Uber's entry increased public transit ridership in large cities, but it decreased transit ridership in small cities. Interestingly, Uber's entry increased the ridership of transit agencies that had below-median public ridership while decreasing ridership for the transit agencies with above-median ridership.

Curtis et al. (2019) investigated the partnership between transit agencies and Transportation Network Companies (TNCs) and the potential opportunities and challenges. TNCs provide transportation via mobile apps that connect riders to available drivers nearby. This also known as ridesharing. They gathered information from 20 transit agencies through a survey and a follow-up interview. They provided comprehensive information about the data in terms of partnership development and implementation policies. The results indicated that motivations for engaging in partnerships were strong when TNCs served as a specific type of service, met or responded to a specific policy goal or challenge, and demonstrated innovation and flexibility to experiment. Moreover, the most common market of the partnership was the first mile/last mile service and customers of ADA paratransit or dial-a-ride (DAR) services. Finally, the most common partnership implementation involved TNC trips subsidized directly by transit agencies.

Table 1 presents a summary of existing studies that looked at the first/last mile connection perspectives of transit usage. The first group of studies focused on the role and potential of TNC services, and the second group of studies looked into the effects of integrating AV technologies into the transit network services.

In general, the literature found positive impacts of integrating ridesharing and AV technologies with transit services, in terms of reducing VMT, reducing driving trips, saving travel time/cost, reducing operational cost, and increasing efficiency. Both substitution and complement effects have been noted in the literature. Most studies employed agent-based modeling and

simulation techniques to investigate the impacts of the integrated system. Some other papers incorporated a mode choice component that was able to capture modal shifts.

Table 1 Summary of Literature

Study	Study Year	Study Area	Method	Service Type	Study Purpose
Davidson et al. (*8*)	2017	New York City, USA	Data comparison	TNC	Relationship between Uber and transit usage, the role of smartphone apps
Alemi and Rodier (*6*)	2018	San Francisco, USA	Agent-based demand and supply simulation	TNC	Benefit evaluation, mode shift from drive alone to BART
Stiglic et al. (*16*)	2018	n/a	Ride-matching algorithms	ridesharing	Benefit evaluation
Yan et al. (*11*)	2018	University of Michigan Ann Arbor campus, USA	Mixed logit model based on SP scenarios	ridesourcing	Evaluate traveler responses to a proposed integrated service
Hall et al. (*12*)	2018	United States	Difference-in-differences approach	TNC	Impacts on transit ridership
Jaller et al. (*14*)	2019	San Francisco, USA	Agent-based model simulation & optimization	ridesharing	Benefit evaluation, mode shift
Curtis et al. ***(13)***	2019	United States	Survey	TNC	investigated the partnership between transit agencies and TNC and the potential opportunities and challenges.
Vakayil et al. (*2*)	2017	Washington DC, USA	Simulation	autonomous mobility-on-demand system	Service integration, transit with AMOD vs. AMOD only, fixed demand,
Shen et al. (*15*)	2017	Singapore	Agent-based Simulation	on-demand SAV	Performance evaluation
Farhan et al. (*3*)	2018	Seattle metropolitan area, USA	Simulation	Shared Autonomous Electric Vehicles (SAEVs)	Operations and performance by vehicle range and charging infrastructure
Pinto et al. (*7*)	2018	Chicago, USA	Integrated mode choice and	SAV	Assess the impacts on transit demand

Study	Study Year	Study Area	Method	Service Type	Study Purpose
			dynamic assignment		
Berrada et al. (*5*)	2019	Palaiseau, France	Simulation with 4-step demand model	autonomous taxi, ridesharing	performance evaluation, aTaxis replacing BRT in connecting transit stations with destinations.

1.3 DATA AND METHODOLOGY

This section describes the data compiled for this study as well as the modeling methodology used to analyze the mode choice of travelers' access and egress segments.

1.3.1 Transit On-Board Survey

A transit on-board survey was implemented by ETC Institute with AECOM, Connetics Transportation Group (CTG), and Resource Systems Group (RSG) on the team for the transit agency in the Orlando metro area, the LYNX. The data collection began in January and ended in April of 2017. The study was conducted in the Orlando metro area, covering the LYNX service area. In total, the survey collected 13,181 responses.

The 2017 LYNX on-board survey recorded the trip information for each respondent, including the trip origin, boarding stop, transfer stop(s), if any, alighting stop, and destination of the bus trip for which the survey took place. The respondent's home address, together with sociodemographic information, were also requested. Trip information such as origin/destination place type, access, and egress modes, boarding and alighting time, fare paid for the trip, etc., were also collected in the survey.

Figure 1 illustrates the locations involved in a transit journey. Note that home is not necessarily the origin of a trip as some respondents took the survey on their way home from work.

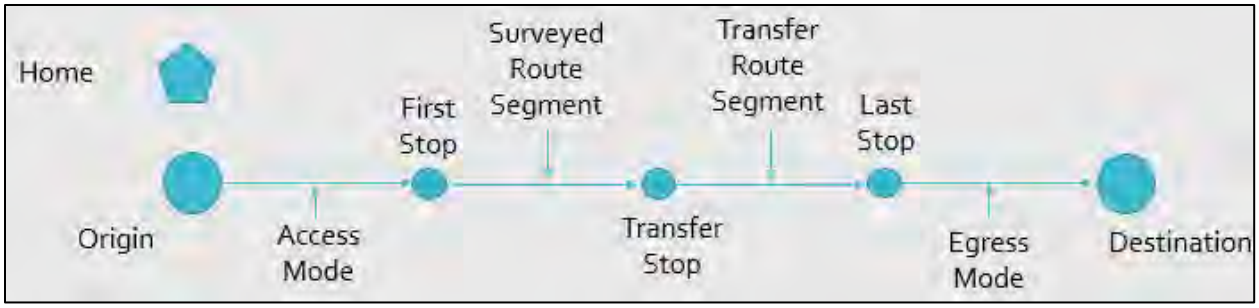


Figure 1 Organization of Locations in a Transit Trip

The longitude and latitude of an origin, destination, or home location was derived from geocoding with the address provided by the respondent. Some respondents did not enter complete addresses (e.g., missing street numbers) for their homes, origins, or destinations. Thus, longitude and latitude of these locations were not accurate and were removed from the analysis of access and egress mode choices.

Figure 2 shows the home locations of the on-board survey respondents. Most of the home locations of the respondents are centered in Orlando and covered by the LYNX and SunRail networks. However, it can be seen that some of the respondents were from out of town for business or sightseeing on the day surveyed (i.e., tourists from out of state are excluded from this study).

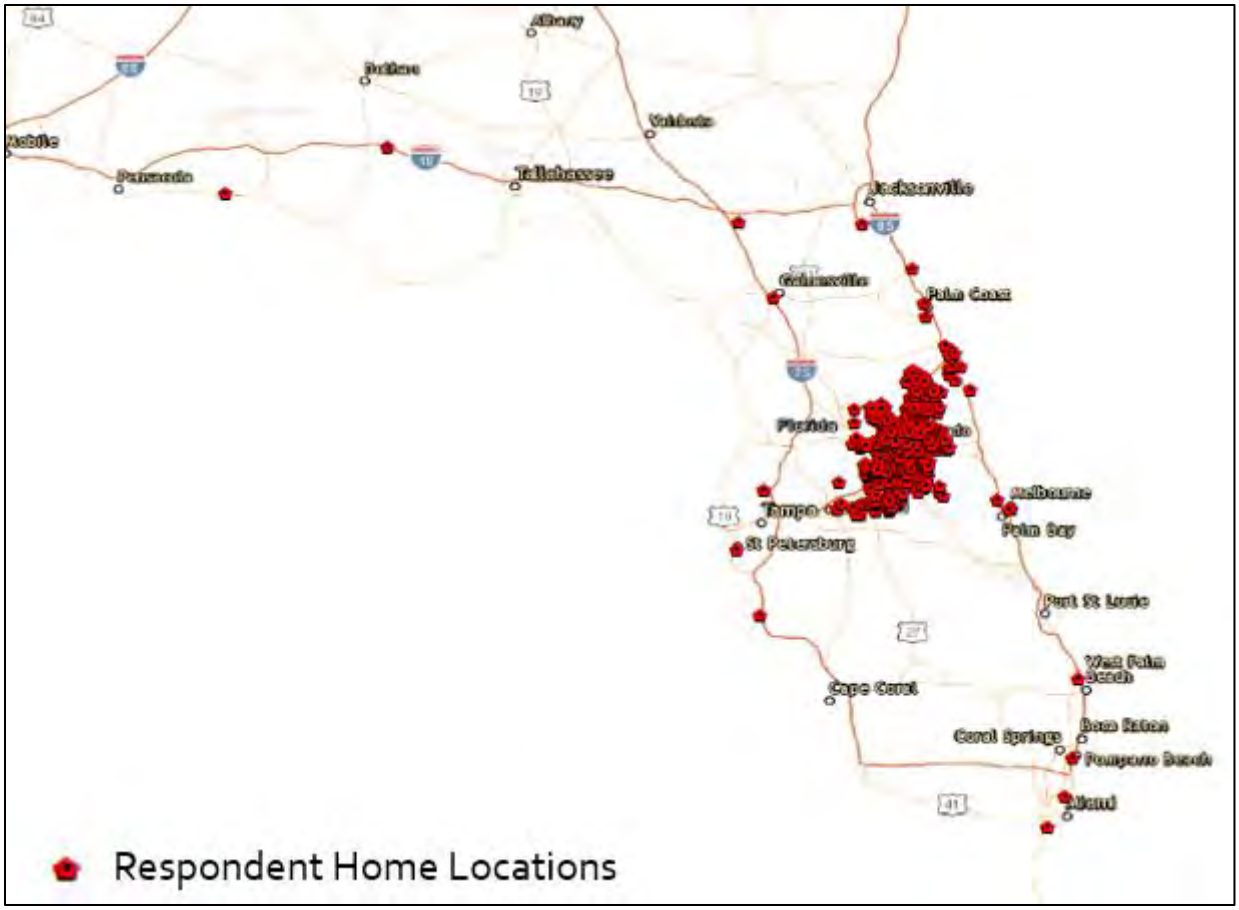


Figure 2 Home Locations of Survey Respondents

Figure 3 shows a closer look for some of the locations of the respondents' trip origins. It can be seen that most of the trip origin locations are next to streets and roads (i.e., TIGER/Line data) as the locations were geocoded by finding the corresponding addresses on the street GIS data.

Figure 4 shows a similar map for the respondents' destination locations.

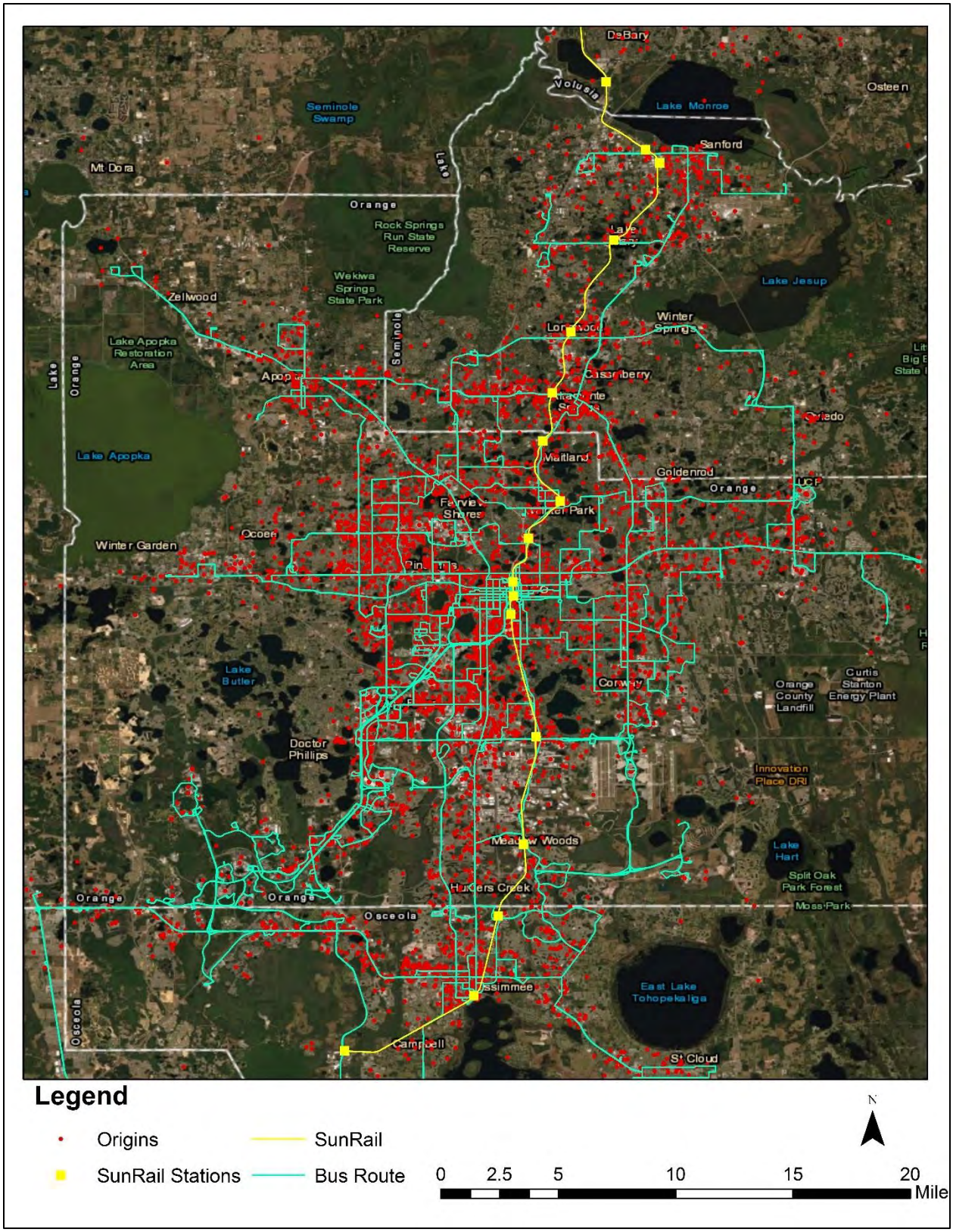


Figure 3 Locations of Trip Origins

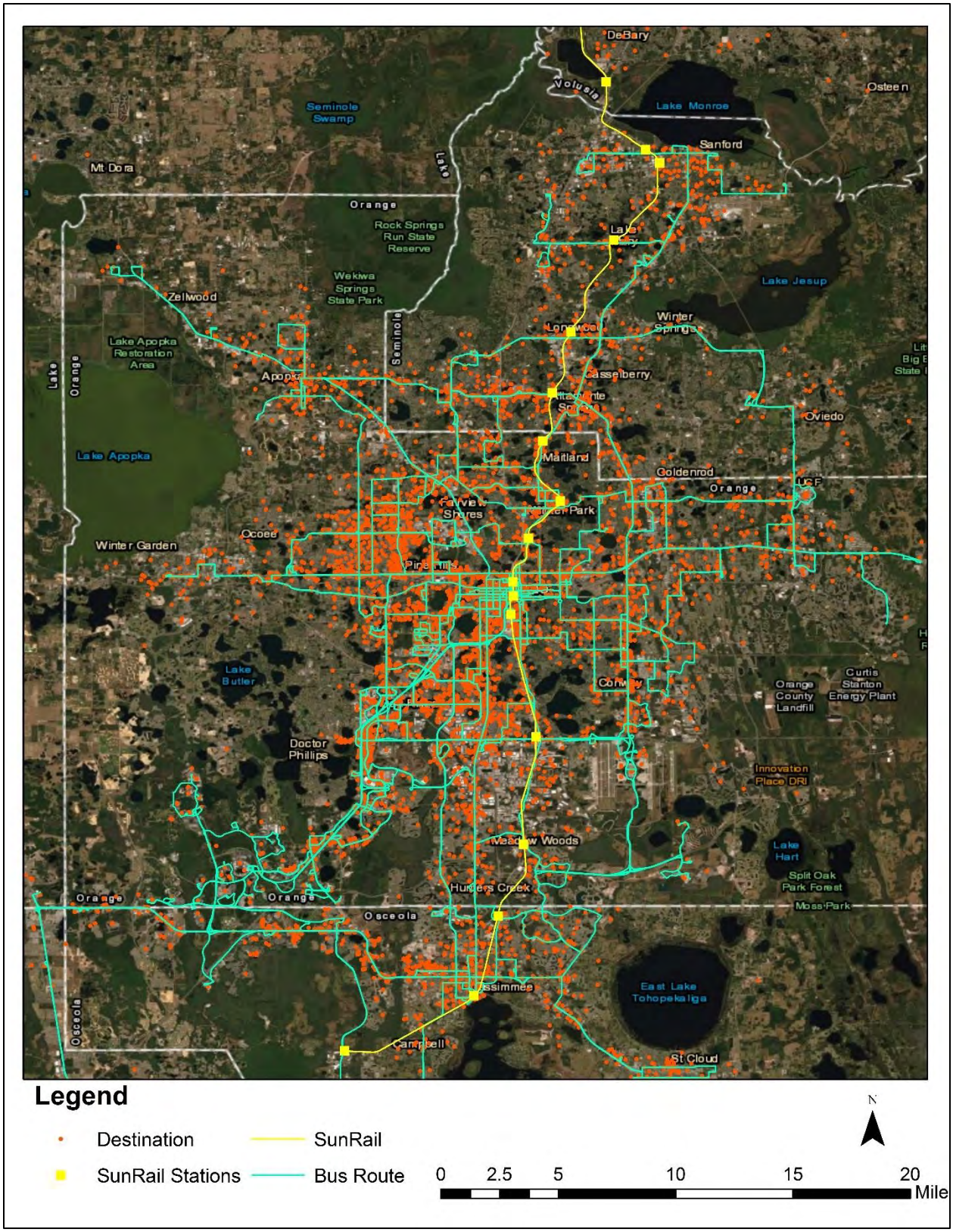


Figure 4 Locations of Trip Destinations

1.3.2 GIS Databases

1.3.2.1 LYNX and SunRail GIS and GTFS Data

To visualize a transit trip from the origin to its destination, the GIS and General Transit Feed Specification (GTFS) data for all LYNX bus routes and bus stops (LYNX, 2020a) were obtained. The GTFS is a data specification for transit route and schedule information (GTFS, 2020). GTFS allows public transit agencies to publish their transit data in a format that can be shared by different software applications. As many transit transfers occurred between LYNX and SunRail (i.e., a commuter rail service serving the area), we also obtained the SunRail links and stations GIS/GTFS data from SunRail (SunRail, 2020).

LYNX provides software application developers with GIS and GTFS data for its bus routes and stops. The bus routes data provide information for each bus route, including route number, name, and frequency of service on weekdays, Saturdays, and Sundays. Bus stops data contain the longitude and latitude, name, address, and amenities information for each bus stop and bus shelter in the LYNX network. In addition to fixed-route services, LYNX also offers an on-demand transit service called NeighborLink (LYNX, 2020b). The service areas of NeighborLink are defined as polygons in LYNX's GIS database. SunRail also offers GTFS and station GIS data for the commuter rail services that include schedule, geocoded station locations, and fare information.

GIS data for the streets and roads were also obtained from US Census's TIGER/Line geodatabase (US Census, 2020) in order to visualize the streets that connect the origins to the boarding bus stops. These data were compiled and integrated into the GIS database.

1.3.2.2 Visualizing Access and Egress Distances

In order to visualize how the access modes vary by distances between the origins and first transit stops, we created a GIS line layer, in which each line connects a pair of origin and the first transit stop. Figure 5 and Figure 6 shows how the access and egress links vary by distance and mode in different areas.

It should be noted that the access/egress links shown in Figure 5 and Figure 6 represent the straight-line distances between the origins/destinations and the corresponding boarding/alighting stops for visualization purposes only. The actual access/egress movements should follow street segments. It can be seen in Figure 5 that most of the motorized modes (i.e., dropped off, Uber, Lyft, taxis, drove alone, and car shares) were used by respondents to access SunRail stations and some big LYNX transfer stations. Walking and bicycling as access modes occurred mostly in areas where more LYNX bus routes are available. Figure 6 shows a similar map for the egress modes.

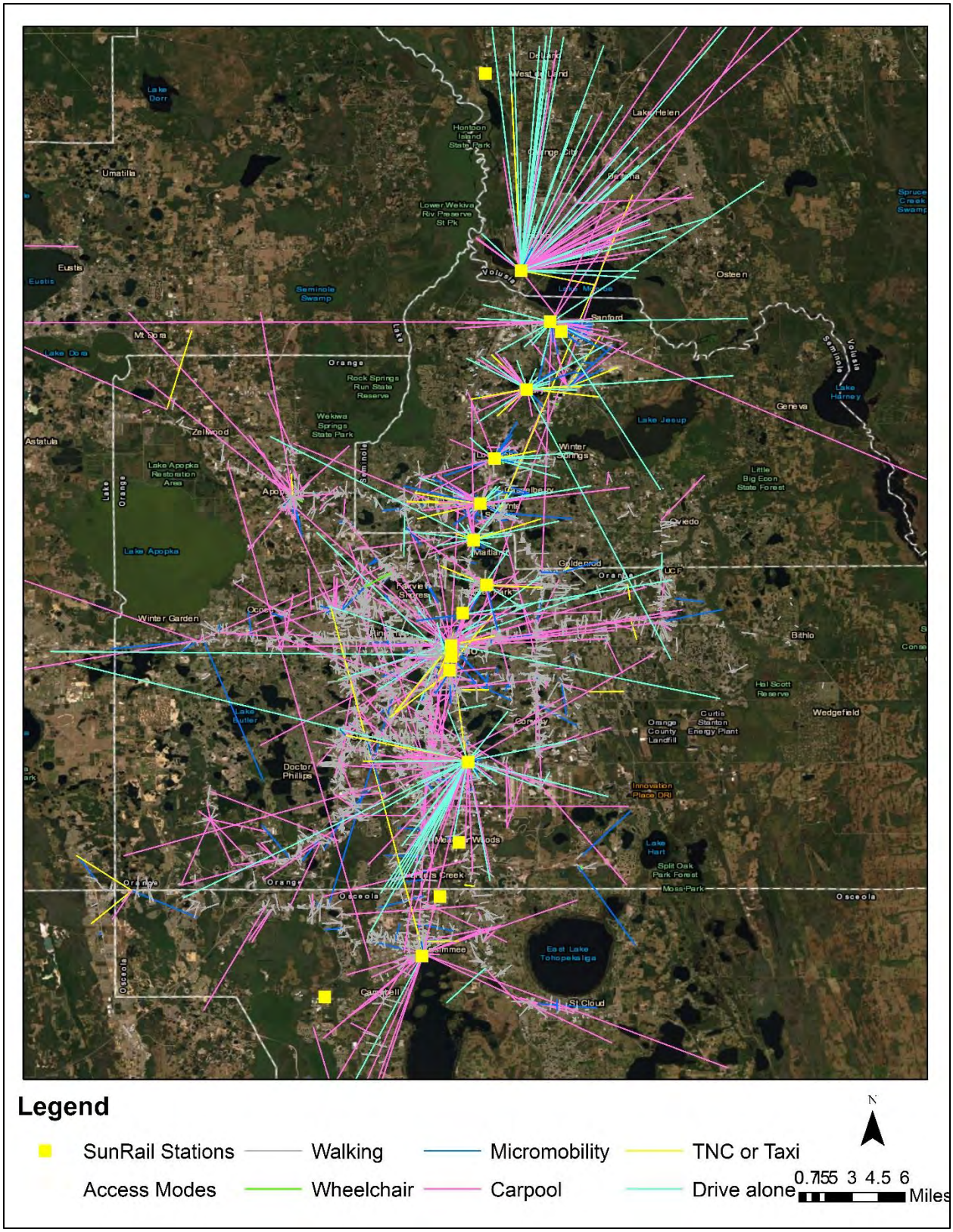


Figure 5 Access Mode Variations (Northern Network)

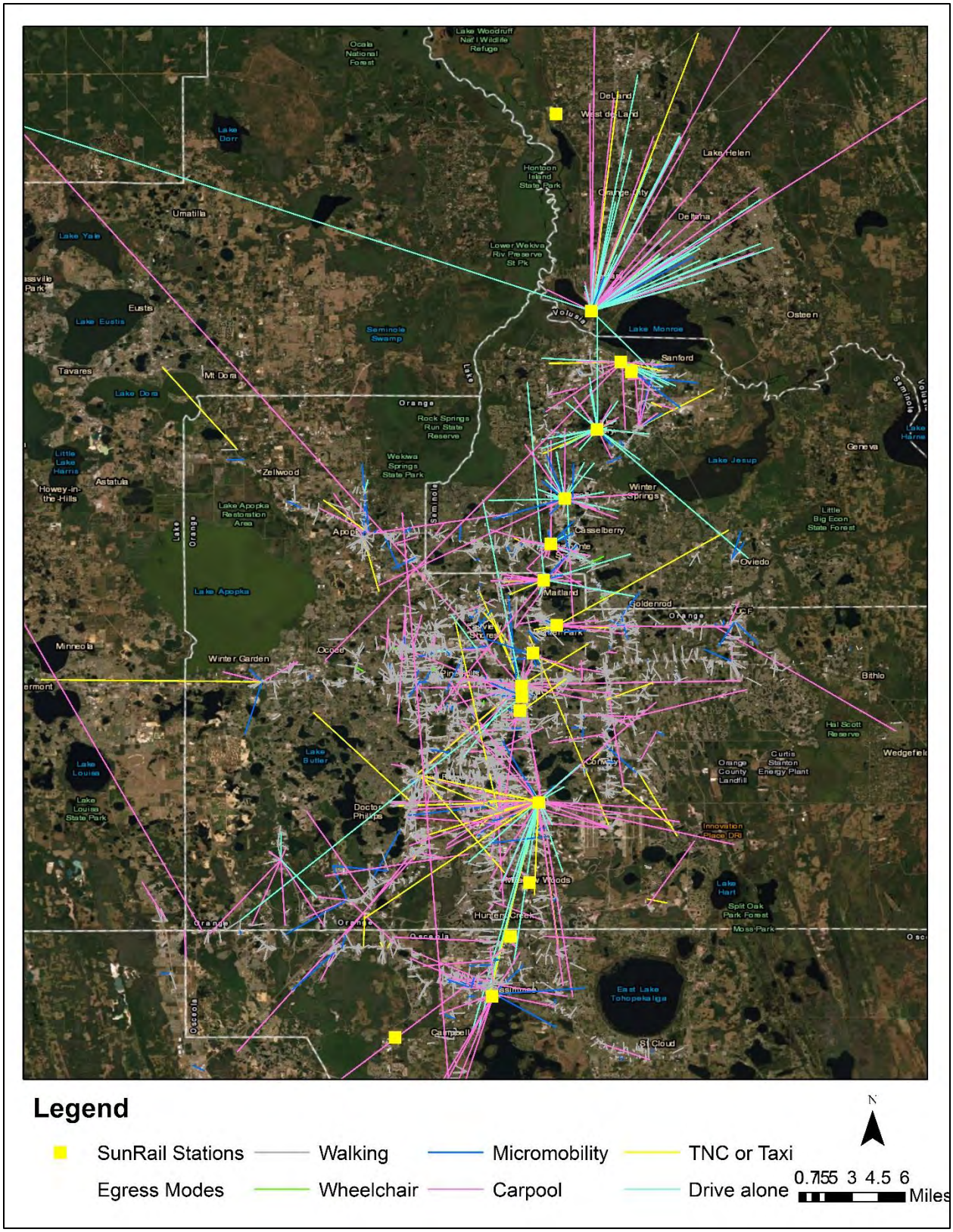


Figure 6 Egress Mode Variations

1.3.2.3 Smart Locations Database

To identify the association between transit use and neighborhood land use characteristics in the study area, we obtained EPA's Smart Locations Database (SLD), which contains data measuring the demographics, socioeconomics, and built environment of census block groups (CBG). Variables for each CBG are divided into seven categories, including demographics, employment, density, diversity, design, transit, and accessibility (EPA, 2014). Demographic variables of the SLD were derived from the 2010 Census data. These include population, households, and household workers by earning levels. Employment variables reflect job activities and workplace-based socioeconomic characteristics in each CBG.

Density variables summarize the numbers of the population, households, and employment within a CBG per unprotected block group acreage, which represents a land area that is not protected from development activity. For CBGs where the unprotected area represented less than one half of one percent of its total area, density metrics were calculated based on total land area rather than the unprotected area.

Diversity variables measure the relative mix of land uses within a CBG. These variables measure the mix of CBG housing unit counts, and employment counts broken down by employment sectors. Because the sizes of CBG vary significantly with respect to urbanization, there is a significant limitation in the accuracy of diversity measured by these metrics. For example, a very large CBG may have a very low-density of diverse land-use activities that are spatially separated within the CBG. Thus, when an individual part of the CBG is examined, its density and diversity are both very low. However, the CBG is regarded as having high diversity. Such a limitation needs to be considered when interpreting the results of analyses involving the diversity variables.

The design variables measure urban design features in terms of street network density and street intersection density by facility types (i.e., automobile, multimodal, or pedestrian). It is important to note that no information regarding the presence or quality of sidewalks or bike paths is included in the SLD. Thus, a CBG with high densities of streets and intersections is not necessarily friendly to walkers or bicyclists. This is another data issue that needs to be considered when analyzing the design variables.

The transit variables measure the availability, proximity, frequency, and density of transit services for each CBG. Data sources for these variables include GTFS data from over 200 transit agencies throughout the United States. Data for transit services with fixed guideways such as rails, streetcars, ferries, and some bus rapid transit routes are also included in the SLD.

The accessibility variables measure the number of jobs and or working-age population accessible within a 45-minute commute via automobile or transit from a CBG. A travel-time decay function is used to weigh jobs and workers by travel time such that activities closer to the origin CBG carry higher weight in accessibility measurement than those that are further away (EPA, 2014).

Table 2 summarizes the variables in the SLD by categories.

Table 2 Summaries of Variables in EPA's Smart Location Database

Categories	**Variables**
Demographics	Housing units (HU), households (occupied housing units), population, percent of the population that is working-aged, number of households by car ownership (zero, one, two or more cars), Percent of households by car ownership (zero, one, two or more cars), number of workers, number of workers by earing levels ($1250/month or less, $1250 - $3333/month, more than $3333/month), percent of low wage workers (earning $1250/month or less).
Employment	Total employment, Retail jobs, Office jobs, Industrial jobs, service jobs, entertainment jobs, education jobs, health care jobs, public administration jobs, number of employees by earing levels ($1250/month or less, $1250 - $3333/month, more than $3333/month), percent of low wage employees (earning $1250/month or less).
Density	Gross residential density (HU/acre), Gross population density (people/acre), Gross total employment density (jobs/acre), Gross retail employment density, Gross office employment density (jobs/acre), gross industrial employment density, gross service employment density, gross entertainment employment density, gross education employment density, gross health care employment density, gross public administration employment density, Gross activity density (employment + HUs).
Diversity	Jobs per household, employment entropy, employment, and household entropy, trip production and attraction equilibrium index, regional diversity, CBG household workers/job, deviation of CBG ratio of household workers/job from the regional average ratio of household workers/job, deviation of CBG ratio of jobs/population from regional average ratio of jobs/population, DBG household/job equilibrium index
Design	Street network density, street network density by transportation mode-orientation (automobile, multimodal, or pedestrian), intersection density, intersection density by transportation model-orientation
Transit	Distance from population-weighted CBG centroid to the nearest transit stop, the proportion of CBG employment within 0.25 or 0.5mile of a fixed-guideway transit stop, aggregate frequency of transit service within 0.25 miles of block group boundary per hour during evening peak period, aggregate frequency of transit service per square mile
Accessibility	Jobs and working-age population within 45 minutes auto travel time, time-decay (network travel time) weighted, jobs and working-age population within 45-minute transit commute, time decay (walk network travel time, GTFS schedules) weighted, proportional accessibility to regional destinations by automobile: employment and working-age population accessibility expressed as a ratio of CBG to total Census MSA (Metropolitan Statistical Area) accessibility, proportional accessibility to regional destinations by transit: employment and working-age population accessibility expressed as a ratio of CBG to total Census MSA accessibility, regional centrality index by automobile: the ratio between a CBG's accessibility score and the maximum accessibility score within a core-based statistical area (CBSA), regional centrality index by transit: the ratio between a CBG's accessibility score and the maximum accessibility score within a core-based statistical area (CBSA)

When combined with the LYNX survey data in a GIS, the SLD variables can be used to identify the association between access or egress modes and neighborhood land use characteristics at the origins or destinations. For example, Figure 7 shows how the access modes vary by the variation of employment to population diversity of the Census block groups. The variable D2R_JOBPOP is calculated based on the total population and total employment of the block group (EPA, 2014). It measures the deviation of a block group's ratio of jobs/population from the regional average ratio jobs/population. It can be seen that motorized modes with longer access distances mostly originated from block groups with lower employment to population ratios, while walk trips with shorter distances occurred in areas with higher ratios.

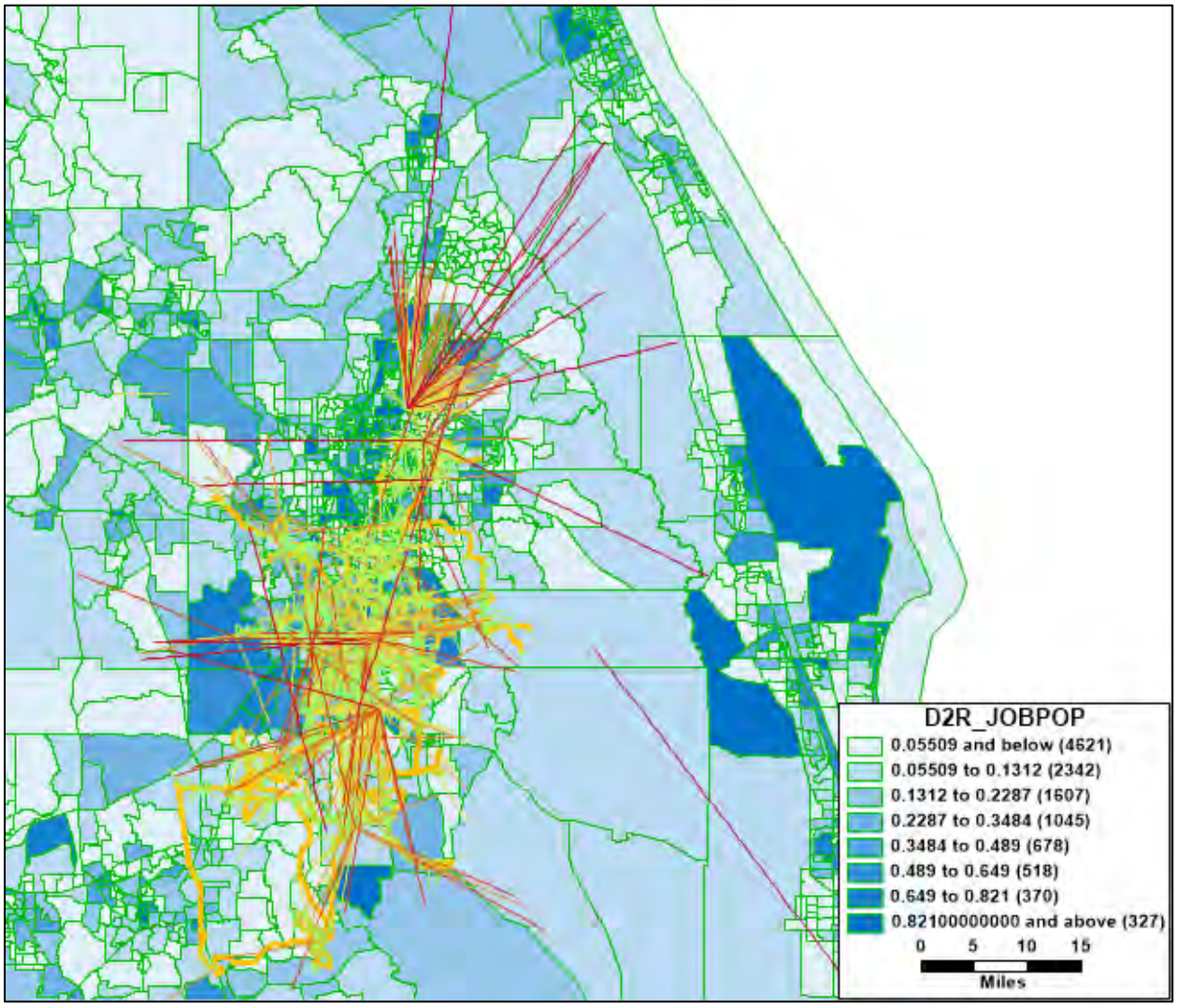


Figure 7 Access Modes Variation and Employment to Population Diversities

1.3.3 Mode Choice Modeling

Two Multinomial logit (MNL) models were developed to analyze the mode choice behavior for access mode and egress mode, respectively. The multinomial logit model is a popular method for exploring the potential relationship between mode choice and the determining factors (Lee et al. 2018). The MNL model has relatively simple mathematical formulation, but at the same time, it accounts for unobserved utilities (McFadden 1973, Ben-Akiva et al. 1985, Koppelman et al. 2000).

The fundamental assumption of the MNL model is that each individual has unobservable, latent utilities for different travel modes, and he chooses the mode that has the highest level of utility (Schwanen and Mokhtarian 2005). The MNL model uses the maximum likelihood method to estimate the impact of explanatory variables on each category of the dependent variable, and it is commonly employed when the dependent variable has more than two categories.

In this model, one category of the dependent variable (usually the one with the highest frequency) is considered as the reference category, and the probability of being in any category of the outcome will be compared to the likelihood of being in the reference category. Therefore, for a dependent variable with M categories, the calculation of M-1 equations is required.

If m=1 is considered as the reference category, then log-odds for choice *m* is as follows:

$$ln\ \frac{P(Y_i = m)}{P(Y_i = 1)} = \alpha_m + \sum_{k=1}^{K} \beta_{mk} X_k = Z_{mi} \qquad (1)$$

Where,

$\alpha_m$is the constant term for the *m*th choice,

β_{mk} is a vector of coefficients for the *m*th choice and the *k*th variable,

X_k is the set of explanatory variables,

Z_{mi} is the predicted log-odds for the *i*th observation

Consequently, there will be M-1 log-odds, one for each category relative to the reference category. When there are more than two categories for the dependent variable, the predicted probability of observation *i* choosing *m*th mode would be calculated as follows (Green 2003):

$$P\ (Y_i = m) = \frac{\exp\ (Z_{mi})}{1 + \sum_{h=2}^{M} (Z_{hi})}, \qquad \forall h = 2, \dots M \qquad (2)$$

1.4 RESULTS

This chapter presents the results in two sections. The first section presents descriptive statistics for the transit trips from the 2017 LYNX survey, including trip characteristics, user characteristics, and spatial characteristics. The second section presents the mode choice model results and elaborates on the impacts of various variables on the individuals' mode choice behavior for access and egress mode.

1.4.1 Transit Travel Patterns

The survey data included 13 access/egress modes, which were aggregated into six major modes for this analysis, as shown in Table 3 below. TNC and Taxi trips were combined into one category, given the very small share of TNC trips, also considering the similarity of the two services. These taxi trips could represent a potential market for TNC services.

Table 3 Mode Category

Mode Group	Origin/Destination Transport Code	Origin/Destination Transport From Survey
Walk	1	Walk
TNC or Taxi	8	Taxi
	9	Uber, Lyft, etc.
Micromobility	2	Personal Bike
	3	Bike share
	12	Skateboard
	13	Scooter
Drive Alone	5	Drove alone and parked
	7	Car share (e.g., Zip Car, etc.)
K&R, Carpool, Shuttle	4	Was dropped off or picked up by someone
	6	Drove or rode with others and parked
	11	Shuttle
Wheelchair	10	Wheelchair

Table 4 illustrates the mode share for access and egress trips based on weighted survey data. Some survey records had missing information for access/egress mode. As shown, walking was the predominant mode for both access (90.5%) and egress (92.4%) trips. The next most popular mode used for access and egress trips was dropped off/picked up, carpool and shuttle. Interestingly, micromobility (Bikeshare, Bike, Scooter, Skateboard) showed a non-trivial market share for both access and egress usage. Only a small portion of trips was made by TNC or taxis for access (0.3%) and egress (0.5%) trips.

Table 4 Distribution of Access/Egress Modes

Mode	Access Trip		Egress Trip	
	Frequency	Percentage	Frequency	Percentage
Walking	83,076	90.5%	84,811	92.4%
TNC or Taxi	295	0.3%	481	0.5%
Micromobility	2,060	2.2%	2,377	2.6%
Drive Alone	1,700	1.9%	1,186	1.3%
K&R, Carpool, Shuttle	4,535	4.9%	2,809	3.1%
Wheelchair	122	0.1%	123	0.1%
Total	91,787	100.0%	97,787	100.0%

1.4.1.1 Trip Characteristics

Figure 8 presents the percentage of individuals that took the same trip in the opposite direction by their access and egress mode group. It shows that a majority of those drove alone users, especially for access to transit, were making round trips, which makes sense as they need to get the car on the way back with the exception of carsharing users who may just return the car at the station. On the other hand, most TNC or taxi users did not have a trip in the opposite direction, which may indicate the flexibility of these modes.

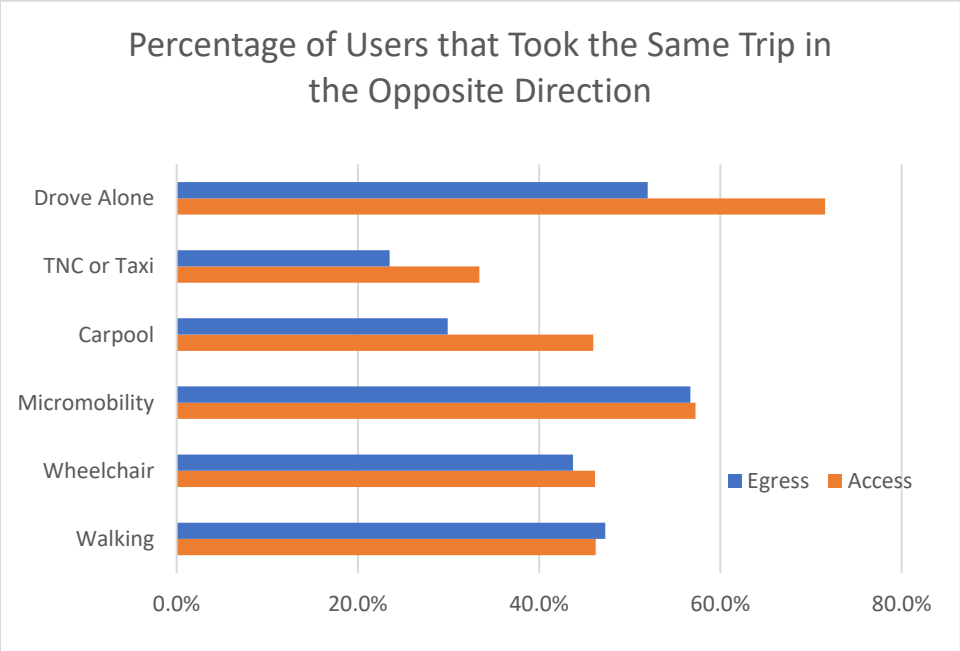
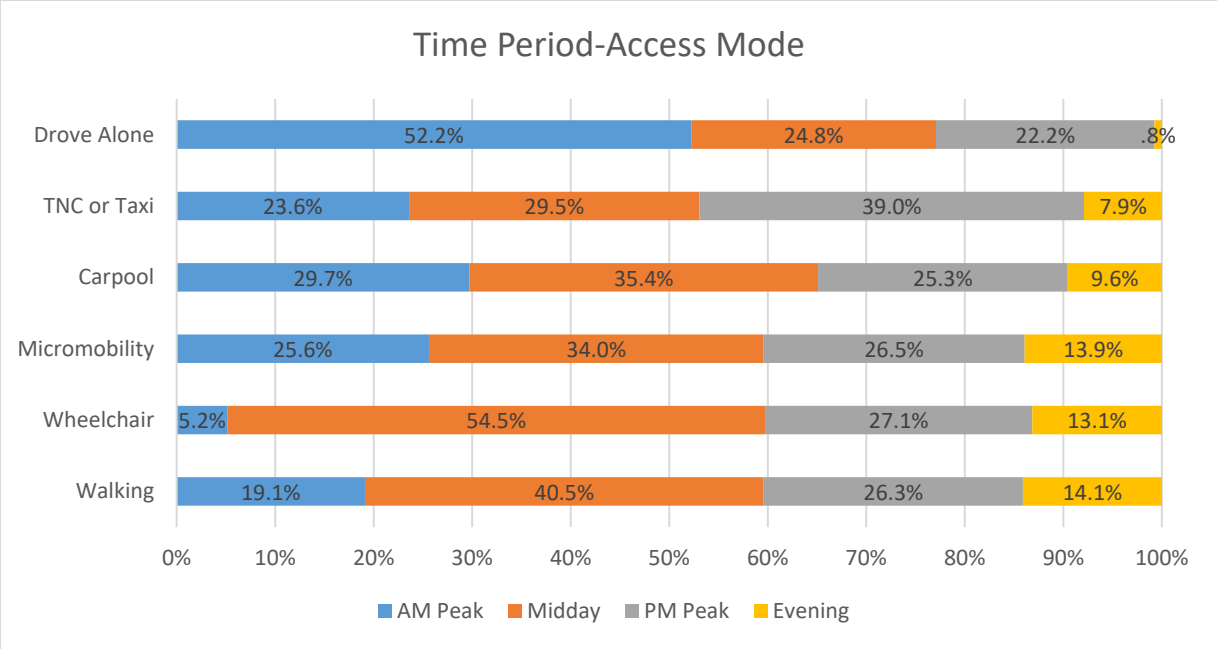


Figure 8 Trip in Opposite Direction by Mode

Looking into the mode groups by time of day, Figure 9 shows very similar patterns for both access and egress modes. Evening trips showed much less share than other time periods, as transit services have limited service hours during the evening period. Compared to other

modes, drove alone trips were more likely to take place during AM peak hours, probably for work purposes. Wheelchair users were more likely to use the transit service during the midday period and much less likely in AM peak hours compared to other users.



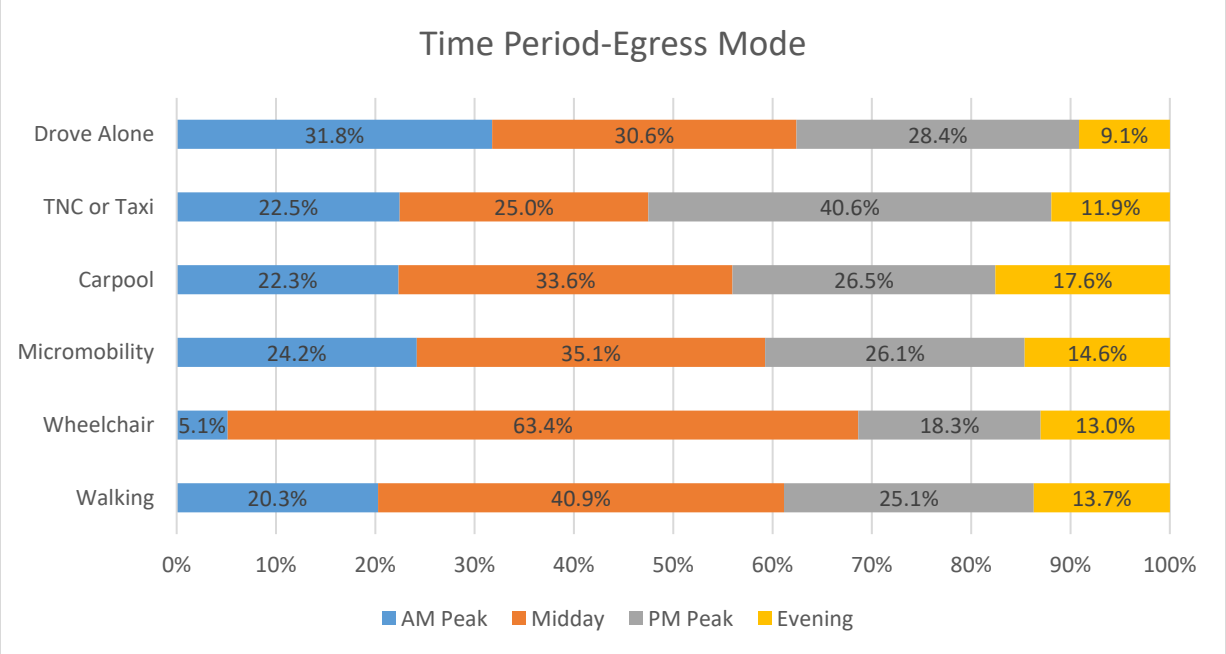


Figure 9 Distribution of Time Period by Mode

To further investigate individuals' trip patterns, Figure 10 presents the distribution of access and egress mode by trip origin place type. Home and work represent a significant portion of trips for all modes, as expected. 90% of drove alone trips originated from home. Comparing

among other place types, wheelchair users showed a significant share of medical trips, while TNC and taxi users showed higher chances of connecting to airports, especially for egress purposes.

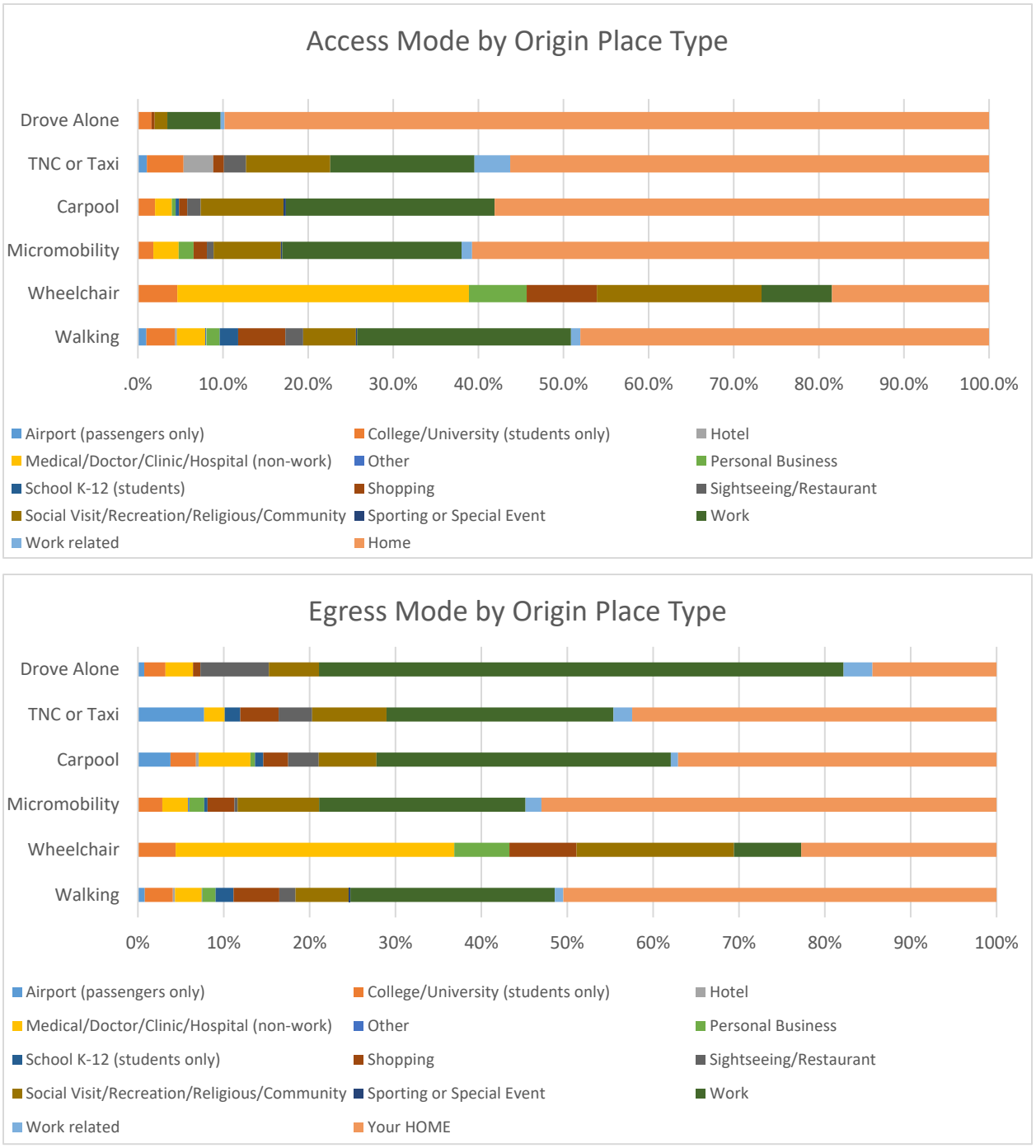


Figure 10 Trip Origin Place by Mode

Similarly, Figure 11 presents the distribution by trip destination place type. Besides home, work, and work-related trips, it shows significant college/university trips taken by TNC or taxies for access purposes. Interestingly, there was significant use of micromobility for college/university trips for both access and egress modes. It also can be seen that more than 50% of carpool trips as egress mode were going home, which probably due to the easiness to prearrange carpool trips and services with fixed destinations.

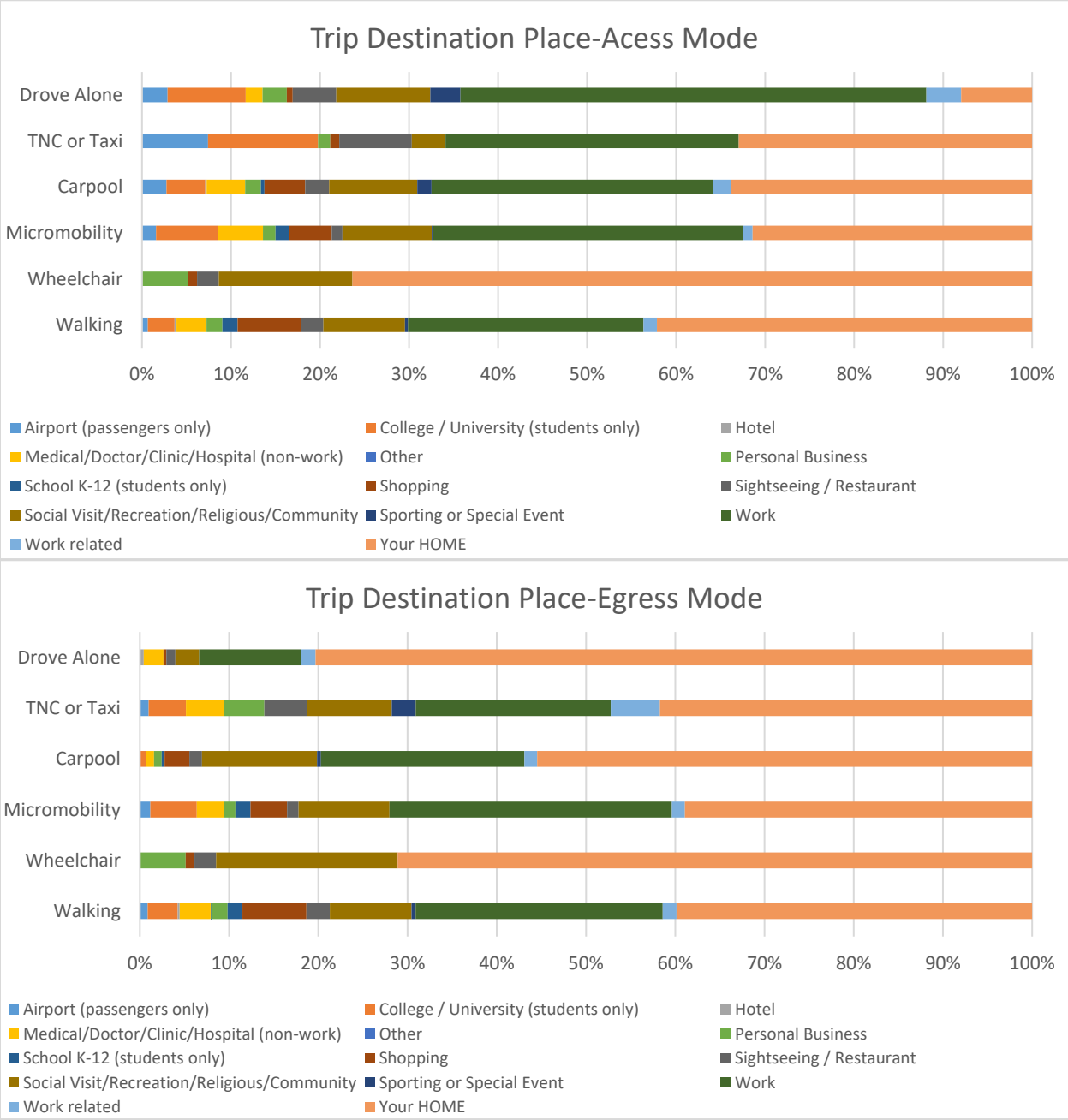


Figure 11 Trip Destination Place by Mode

1.4.1.2 User Characteristics

This section focuses on the patterns in user characteristics. Figure 12 shows the usage of access and egress modes by gender. It shows that males and females were almost equally likely to use different modes, with the exception of micromobility. A majority of the micromobility users were male for both access (88.4%) and egress (82.5%) purposes. Relatively speaking, female users were less likely to use cars or TNC or taxi for both access and egress trips.

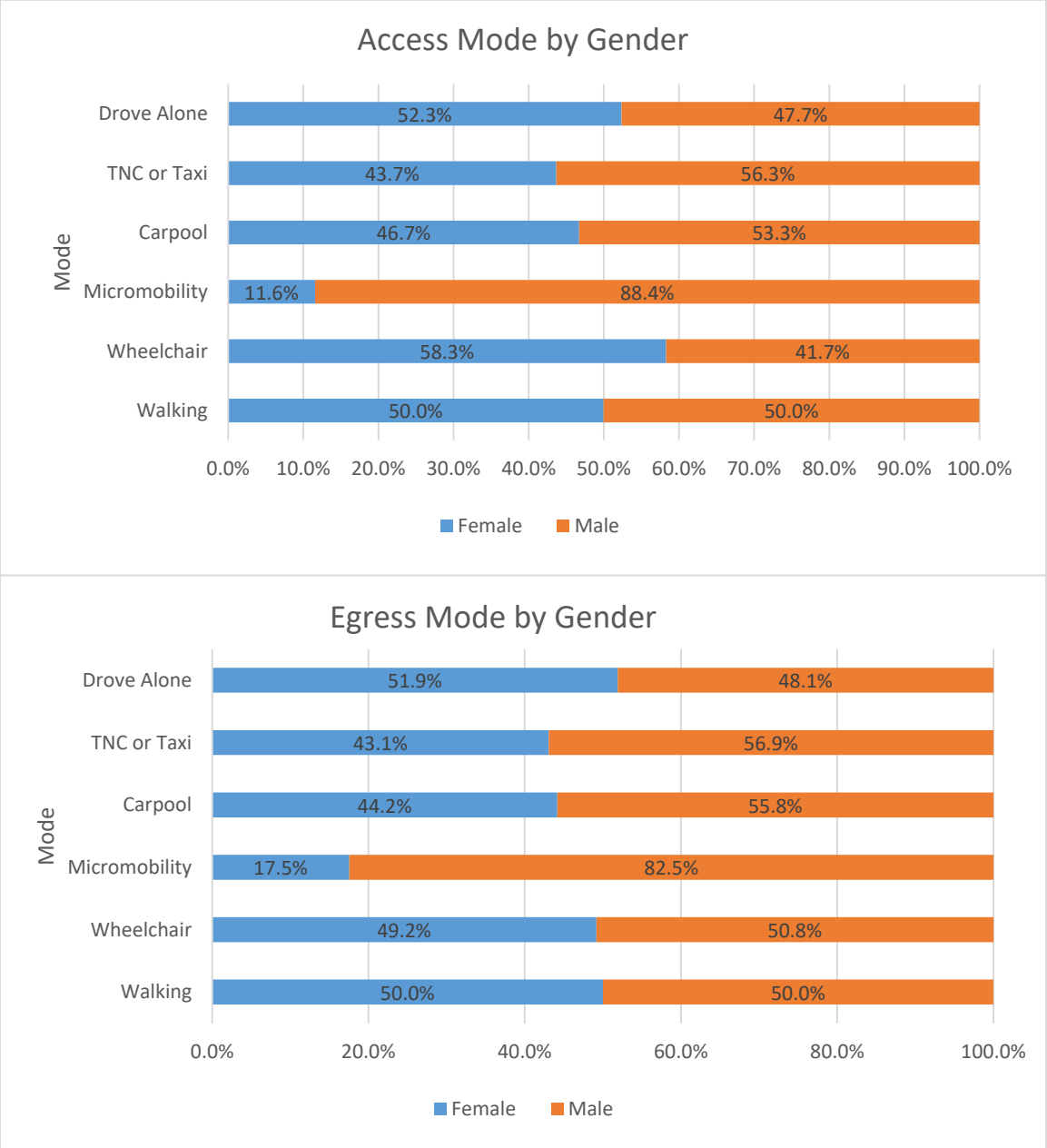


Figure 12 Distribution of Gender by Mode

Figure 13 shows the age distribution for different rider types. Compared with other modes, young adults (18-34 years old) showed a higher propensity to use walk and carpool for access, and walk and TNC or taxi for egress. Middle-aged adults were less likely to walk and carpool for both access and egress. Older adults (above 55) showed high shares of using wheelchairs, for those who did travel using transit services. Interestingly, they were also less likely to use TNC or taxi services for connecting with transit.

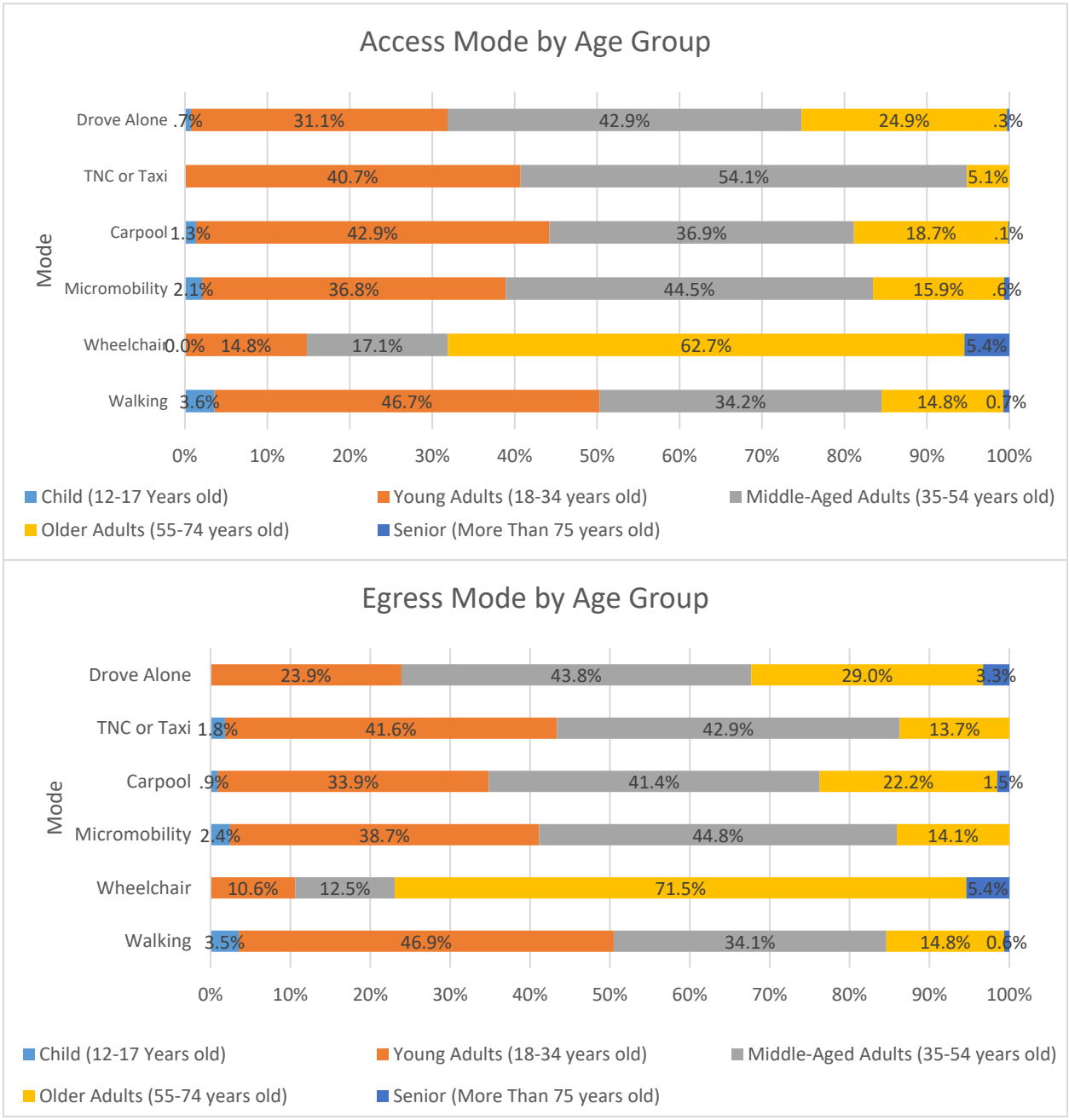


Figure 13 Distribution of Age by Mode

In terms of household income distribution among different user groups, the graphs are presented in Figure 14. As it moves from walking to drive alone mode, there were lower shares of low-income households and higher shares of higher-income households. The pattern is especially clear for access trips. Only 1.3% of the walk access trips and walk egress trips were made by users with very high household income (more than $100K). Comparing between access and egress trips, it seems that the impacts of income on access mode choice were more prevalent than that on the egress mode.

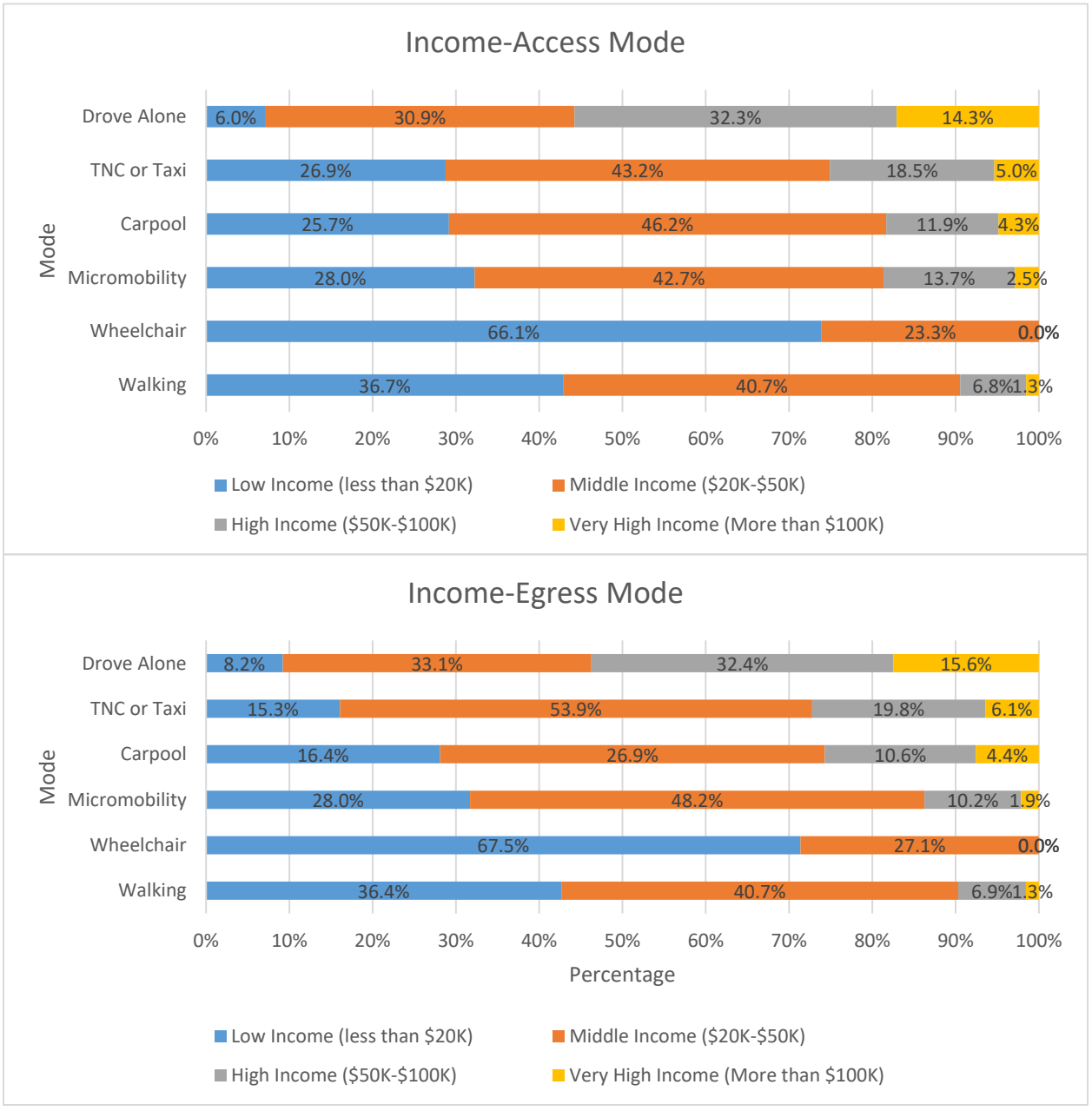


Figure 14 Distribution of Income by Mode

1.4.1.3 Spatial Characteristics

Figure 15 presents the mode share by trip length for the access and egress segments, respectively. As expected, for segments less than 1 mile, walking had much higher shares among the modes than for the longer segments for both access and egress purposes. Generally, as the access length increased, the share of walking mode decreased, and the share of drive alone mode increased. Carpool seems to be the most desirable for trips between 1 to 5 miles, especially for access purposes. Beyond 20 miles, drive alone, and carpool were the most desirable modes, while other modes were unlikely to be used.

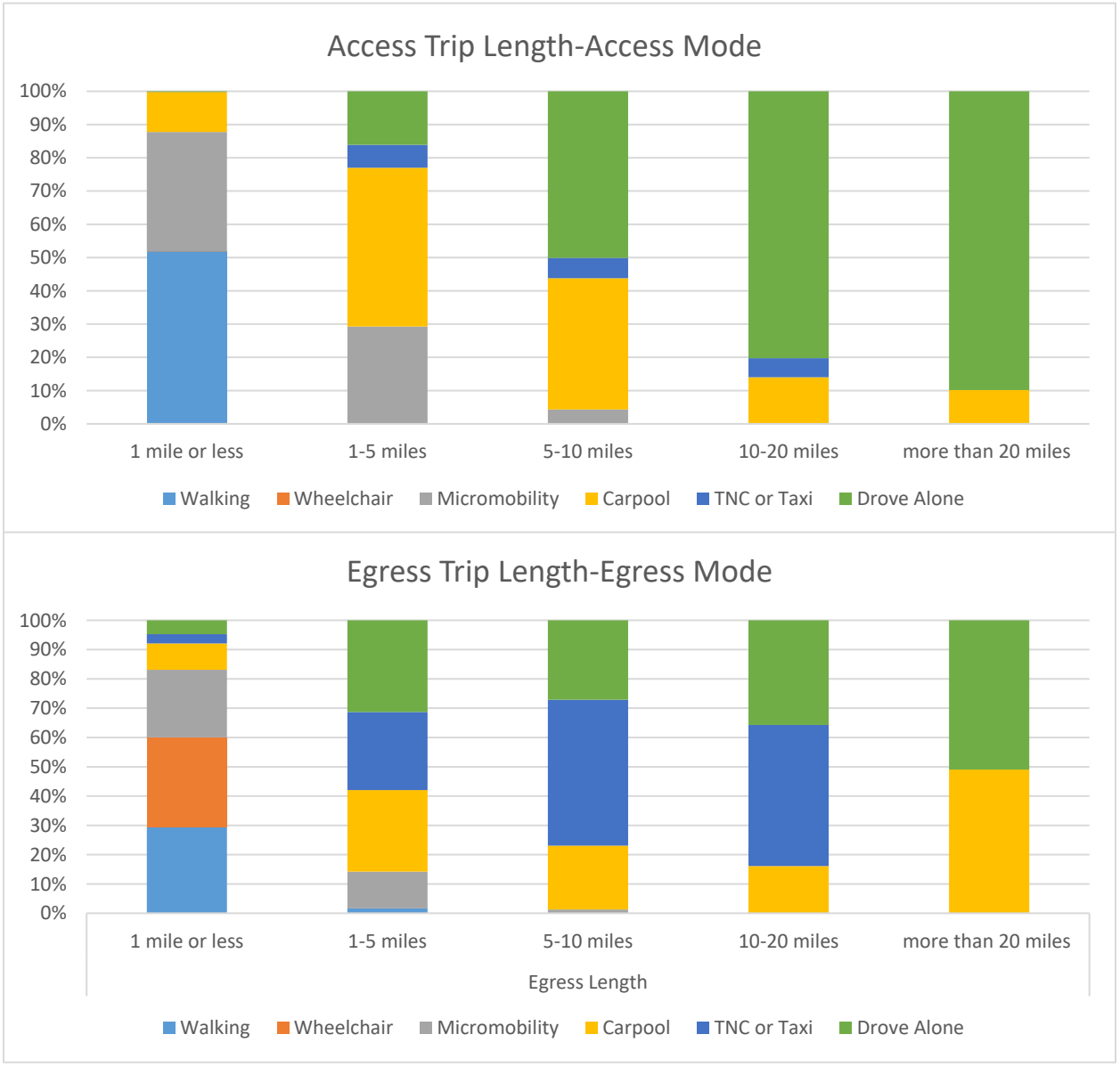


Figure 15 Access/Egress Length by Mode

In terms of the impacts of land use pattern, Figure 16 presents the mode share by urban type, specifically origin urban type for access trips and destination urban type for egress trips. It shows that 22% of TNC/taxi access trips originated in rural areas, a much higher proportion of rural trips than other modes. This indicates a large potential of TNC to connect transit services in low density areas. Carpool also showed a significant share in connecting transit services to destinations in rural areas.

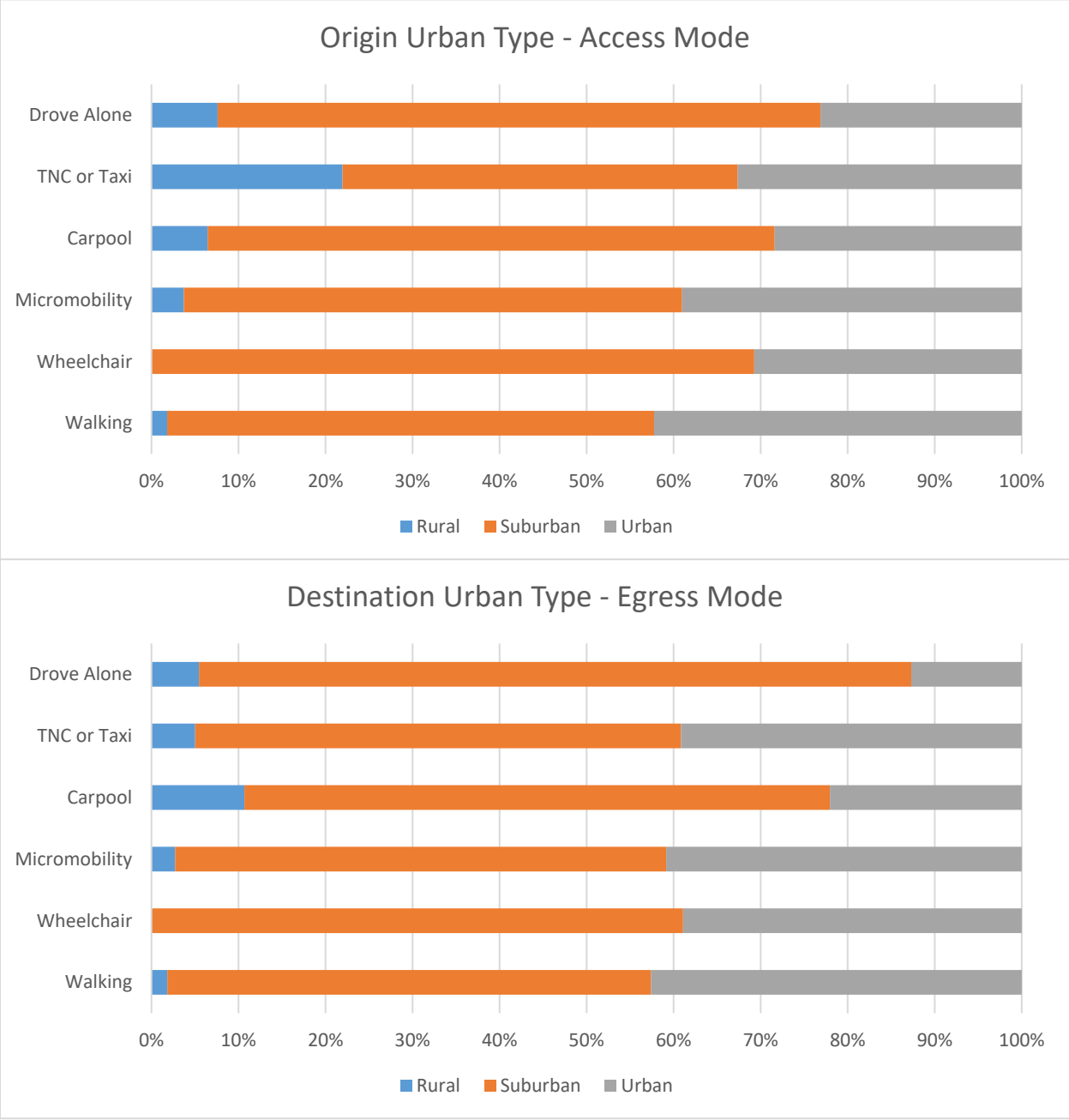


Figure 16 Mode Share by Urban Type

1.4.2 Mode Choice Model Results

As mentioned, to investigate the factors that impact individuals' choices behavior for access and egress trips, two separate multinomial logit models were developed. Walking was considered as the base category in the models. The model results are discussed from three perspectives: trip characteristics, user characteristics, and land-use patterns.

The model performance results for both models are presented in Table 5. Akaike's information criterion (AIC) is a fined method based on in-sample fit to estimate the likelihood of a model for predicting future values (Akaike 1974, Mohammed et al. 2015). Schwarz Criterion (SC) or Bayesian information criterion (BIC) estimates the trade-off between model fit and complexity of the model (Stone, 1979). A lower AIC or BIC value is preferred. As shown in Table 5, both access and egress models showed a lower AIC and BIC value in the full finalized model than the initial model; therefore, indicating acceptable performance.

Another criterion that compares the performance of statistical models is the log-likelihood Ratio Test (LRT). In this study, the LRT detects whether the improvement in the performance of the full model compared to the initial model is significant or not. The difference in the log-likelihood scores for the two models is calculated as:

$$LL = -2[LogL(\beta_{full\ model}) - LogL(\beta_{initial\ model})] \quad (3)$$

If the value of LL is higher than the value of χ^2_{DF} (a chi-square distributed statistic with degrees of freedom equal to the difference in the number of estimated parameters for the two models), It can be stated that the performance of the full model is significantly better than the initial model (Washington et al. 2011).

Again, both full models showed better goodness-of-fit compared to the initial model, and this stands true at a 5% significance level.

Table 5 Model Performance Results

	Access Model		**Egress Model**	
Criterion	Initial Model	Full Model	Initial Model	Full Model
AIC	11512.294	9914.894	9707.009	6528.688
SC	11549.645	11857.156	9744.36	7537.171
-2 LOGL	11502.294	9394.894	9697.009	6258.688
LL	2107.4004		3438.3205	
Max-Rescaled R-Square	0.255		0.442	

1.4.2.1 Access Mode Choice Model

Impacts of Trip Characteristics

Table 6 presents the access mode choice model results for trip-related variables. For a unit change in the explanatory variable, the logit of each mode (which is in log-odds units) relative to the reference mode (walking) changes by its respective parameter estimate, given that the other variables are held constant. The values in the parenthesis represent the standard deviations for the corresponding factor.

Table 6 Model Results for Trip Characteristics

Base Category	Parameter	Wheelchair	Micro-mobility	Carpool	TNC or Taxi	Drove Alone
	Intercept	-6.5879 (0.0023)	-3.6756 (0.0012)	-3.7920 (0.0019)	-5.2544 (0.0033)	-2.9104 (0.0011)
Access Length	**Distance from the origin to transit stop (mile)**		1.3802 (0.0004)	1.0966 (0.0004)	1.2077 (0.0009)	0.8382 (0.0011)
Destination Place	**Airport**	-0.6220 (0.0765)	0.6333 (0.0141)	1.6289 (0.0077)	4.5079 (0.0080)	0.7242 (0.0208)
	Medical, Hospital	-1.7809 (0.0145)	0.1199 (0.0061)	0.4373 (0.0057)	-0.7863 (0.0345)	0.1350 (0.0102)
	Sporting Events	-0.9363 (0.0944)	0.9780 (0.0144)	3.4428 (0.0080)	-3.0877 (0.4182)	2.3677 (0.0155)
	University/college	-0.3235 (0.0214)	0.7579 (0.0050)	0.1250 (0.0060)	1.0043 (0.0122)	1.0782 (0.0064)
Origin Place	**Shopping**	-0.3878 (0.0088)	-0.5704 (0.0085)	-0.6115 (0.0071)	0.3203 (0.0171)	-0.3722 (0.0111)
	Social Visit	0.7676 (0.0057)	0.2216 (0.0048)	0.4953 (0.0041)	0.2782 (0.0125)	-0.2334 (0.0104)
Transfer	**Number of transfers from origin**	-0.1726 (0.0033)	-0.2860 (0.0021)	-0.2565 (0.0019)	-0.2274 (0.0060)	-0.4630 (0.0036)
	Number of transfers to destination	-0.1730 (0.0035)	-0.3844 (0.0022)	-0.2476 (0.0019)	-0.1449 (0.0052)	-0.4698 (0.0035)
Two-way trip	**Trip in the Opposite Direction-Yes**	0.0474 (0.0035)	0.1887 (0.0016)	0.0287 (0.0016)	-0.3186 (0.0053)	0.2456 (0.0024)
Visitor	**Visitor-Yes**	-0.6948 (0.0138)	-0.8573 (0.0061)	0.0533 (0.0039)	-0.8857 (0.0102)	-0.4065 (0.0081)
Time Period	**Midday**	0.0369 (0.0033)	-0.1284 (0.0021)	-0.2435 (0.0020)	-0.3179 (0.0061	-0.2294 (0.0033)
	Evening	-0.2170 (0.0071)	-0.0934 (0.0035)	-0.2140 (0.0033)	-0.7531 (0.0114)	-0.6177 (0.0069)
Month (reference - January)	**December**	0.4852 (0.0070)	0.2096 (0.0039)	0.0561 (0.0038)	-0.1235 (0.0169)	-0.2026 (0.0081)
	February	0.4805 (0.0035)	-0.1917 (0.0024)	-0.1677 (0.0023)	-0.0660 (0.0081)	0.1235 (0.0036)
	November	-0.4131 (0.0284)	0.1986 (0.0137)	0.1971 (0.0132)	0.9348 (0.0364)	-0.1984 (0.0307)

As expected, the longer the access length (distance from the origin to the transit station), the more likely that people would choose micromobility, TNC or taxi, carpool, and drove alone mode over walking.

Those that went to the airport through transit services were more likely to use TNC or taxi, followed by carpool and drive alone modes for access to transit. On the other hand, medical trips and sporting events were less likely to start with TNC or taxi trips for accessing transit. When people were leaving shopping places, TNC and taxi became highly desirable for connecting their transit trips.

The number of transfers showed negative impacts on the probability of choosing any of the modes compared to walking. This may be an indication that those who took transit services despite the inconvenience of transfers might not have the option to use other modes for cost considerations or other reasons. Wheelchairs were more likely to take place in the midday period, while evening trips were likely to be taken by walking compared to AM and PM peak periods, probably most likely for non-work trips.

Figure 17 presents the model results of trip characteristics in terms of their impacts on the use of access modes. The largest positive and negative effects on the use of TNC or taxi was going to an airport trip, and sporting event, respectively. Micromobility had the highest impacts compared to the access length and going to sports or college/universities. Sports events also showed high positive impacts on the use of carpool and drove alone modes.

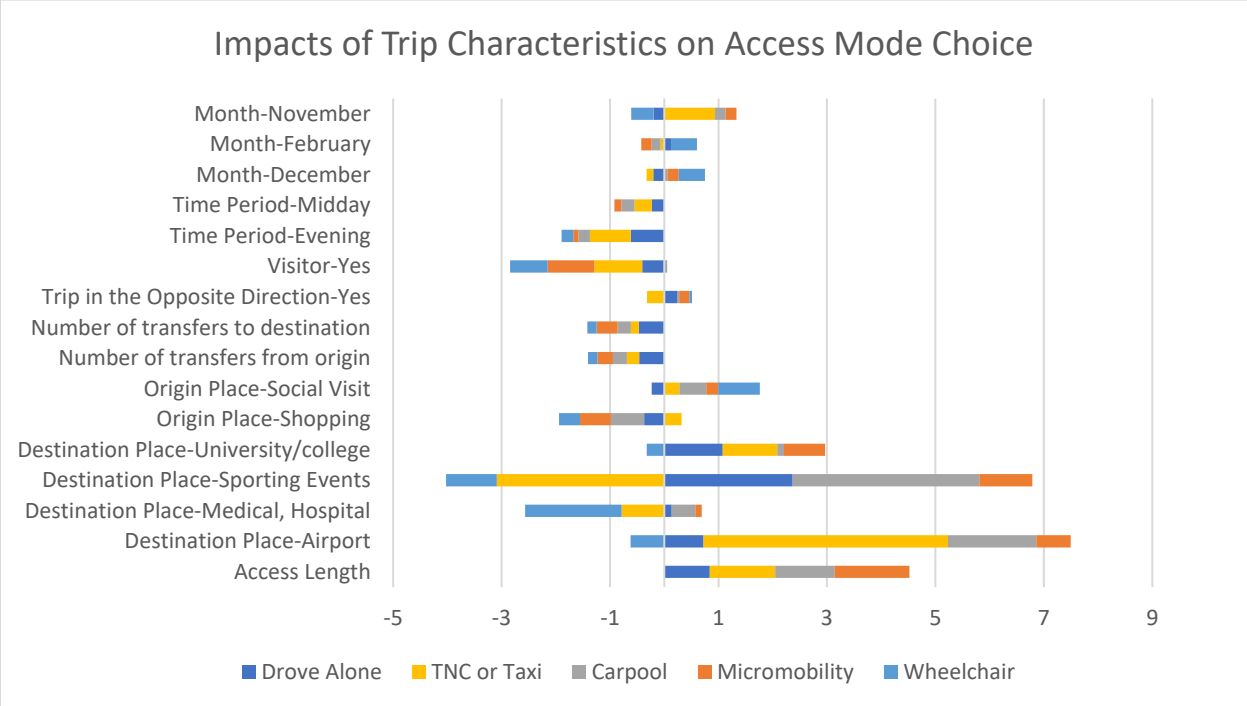


Figure 17 Relative Impacts of Trip Characteristics

Impacts of User Characteristics

Table 7 presents the impacts of user characteristics on the choice of access mode. Interestingly, young adults (18-34 years old) showed a negative association with all modes, suggesting that they were more likely to walk to the transit station. As expected, individuals with disabilities showed the highest positive impacts on the wheelchair, and the highest negative associations with TNC or taxi. This may be because most TNC or Taxi services were not fully compatible with the needs of these users. Households with 4 to 7 vehicles were more likely to choose carpool, drive alone, and micromobility over walking.

As expected, individuals with a driver's license were likely to drive or carpool to use transit services. In view of ethnicity, African Americans and Asians were more likely to use wheelchairs and walking than other modes. Hispanics showed a higher propensity of using TNC or taxi, followed by walking. Male users were more likely to use micromobility modes than females.

In terms of household income, middle-income ($20K-$50K) users were more likely to choose micromobility, followed by driving or carpooling, whereas they were less willing to use TNC or taxi, probably due to cost associated with these modes. Users with high and very high household income showed a high inclination to drive or carpool to access transit services.

Table 7 Model Results for User Characteristics

Base Category	Parameter	Wheelchair	Micro-mobility	Carpool	TNC or Taxi	Drove Alone
	Intercept	-6.5879 (0.0023)	-3.6756 (0.0012)	-2.9104 (0.0011)	-5.2544 (0.0033)	-3.7920 (0.0019)
Age	**Young Adults (18-34 years old)**	-0.2057 (0.0054)	-0.2083 (0.0019)	-0.1087 (0.0017)	-0.2024 (0.0052)	-0.2400 (0.0029)
Disability	**Disability-Yes**	5.6846 (0.0024)	-0.1091 (0.0035)	-0.1853 (0.0040)	-0.4964 (0.0146)	-0.1382 (0.0068)
Driver License	**Driver License-Yes**	-0.3566 (0.0043)	-0.1025 (0.0017)	0.1838 (0.0015)	0.2740 (0.0039)	0.7187 (0.0022)
Ethnicity	**African American**	0.0350 (0.0034)	-0.4618 (0.0020)	-0.1985 (0.0018)	-0.3017 (0.0058)	-0.4848 (0.0033)
	Asian	1.0739 (0.0166)	-0.8116 (0.0132)	-0.1800 (0.0085)	-0.3127 (0.0220)	-0.1952 (0.0135)
	Hispanic	-0.4542 (0.0063)	-0.4058 (0.0025)	-0.1067 (0.0023)	0.1160 (0.0059)	-0.2807 (0.0038)
Gender	**Male**	-0.1961 (0.0032	1.1017 (0.0015)	-0.0648 (0.0016)	-0.3152 (0.0051)	-0.1626 (0.0027)
Number of Vehicles	**Number of Vehicles (4-7)**	-0.3735 (0.0293)	0.2450 (0.0082)	0.7470 (0.0062)		0.6076 (0.0104)
HH Income	**Middle Income ($20K-$50K)**	-0.4078 (0.0048)	0.2643 (0.0018)	0.1236 (0.0018)	-0.1562 (0.0051)	0.1743 (0.0030)
	High Income ($50K-$100K)	-0.5003 (0.0142)	0.4240 (0.0033)	0.7081 (0.0029)	0.4793 (0.0075)	1.3374 (0.0035)
	Very High Income (More than $100K)	-1.0835 (0.0639)	1.8597 (0.0067)	4.7172 (0.0037)	0.9611 (0.0317)	2.6380 (0.0092)

Figure 18 illustrates the impacts of user characteristics on the use of different modes. It shows that income and disability status had the most significant impacts on the choice of access mode. The use of micromobility was mostly impacted by gender and ethnicity.

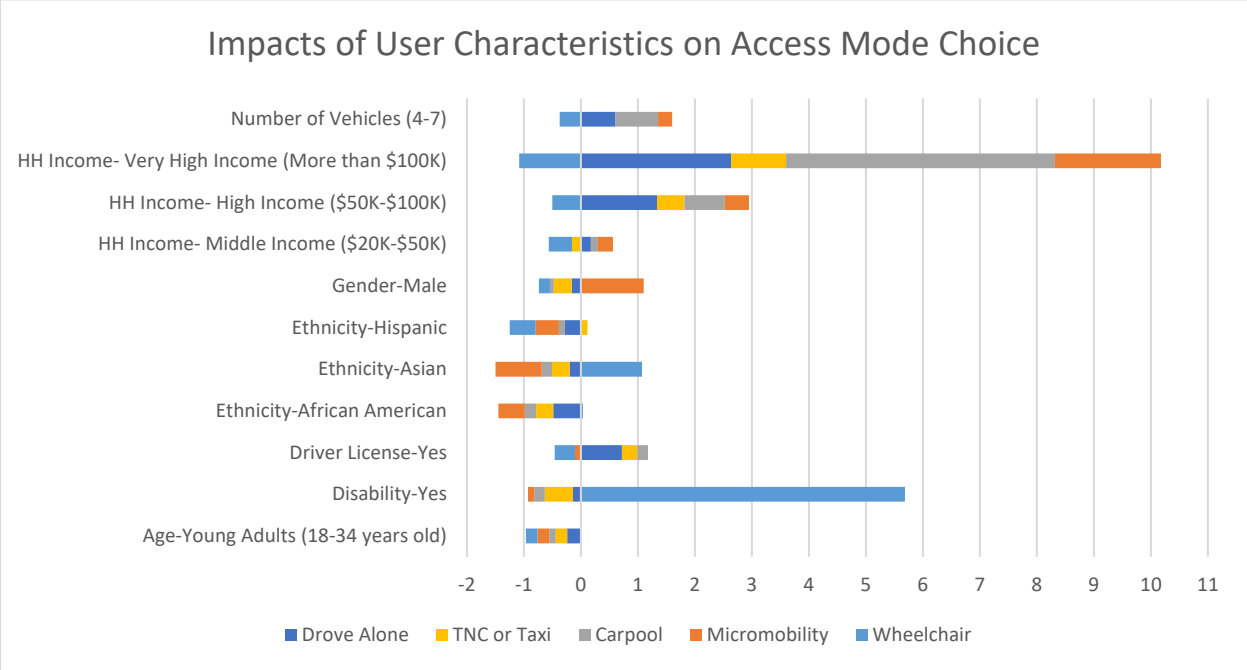


Figure 18 Relative Impacts of User Characteristics

Impacts of Land Use Characteristics

Table 8 summarizes the model results for land-use variables. For the access model, we used variables describing the land use patterns for the origin at the census block group level. As it shows, higher employment and household entropy showed a positive impact on the use of micromobility modes and reduced the probability of using other non-walk modes.

In terms of employment density, education employment seemed to be associated with higher usage of micromobility, TNC or taxi, drove alone, and carpool modes. Interestingly entertainment employment led to higher use of TNC or taxi, wheelchair, and walking than other modes. Industrial employment showed negative impacts on the use of carpool and drive alone modes, while office employment was associated with higher use of micromobility and carpool. High residential density seemed to discourage the use of all modes but walking, while intersection density showed positive impacts on the use of micromobility modes. Surprisingly, the employment rate showed negative impacts on the use of all modes compared to walking except for the wheelchair.

Figure 19 presents the impacts of land-use patterns. Employment and household entropy and regional diversity had the most influential impacts among the land-use variables. Employment and household density had the highest negative associations with the use of TNC or taxi and drive alone mode. Regional diversity had the highest negative impact on micromobility modes.

Table 8 Model Results for Land-Use Patterns

Parameter	Wheelchair	Micro-mobility	Carpool	TNC or Taxi	Drove Alone
Intercept	-6.5879 (0.0023)	-3.6756 (0.0012)	-2.9104 (0.0011)	-5.2544 (0.0033)	-3.7920 (0.0019)
Employment and household entropy	-0.2846 (0.0036)	0.0218 (0.0020)	-0.0740 (0.0018)	-0.3697 (0.0054)	-0.1725 (0.0030)
Gross education(8-tier) employment density (jobs/acre) on unprotected land	-0.0089 (0.0002)	0.0059 (0.0001)	0.0032 (0.0002)	0.0051 (0.0003)	0.0046 (0.0003)
Gross entertainment (5-tier) employment density (jobs/acre) on unprotected land	0.0875 (0.0009)	-0.0246 (0.0007)	-0.0387 (0.0006)	0.0220 (0.0015)	-0.0578 (0.0012)
Gross industrial (5-tier) employment density (jobs/acre) on unprotected land	0.0588 (0.0012)	0.0728 (0.0009)	-0.0349 (0.0008)	0.0837 (0.0025)	-0.0138 (0.0015)
Gross office (8-tier) employment density (jobs/acre) on unprotected land	-0.0479 (0.0007)	0.0431 (0.0006)	0.0104 (0.0004)		-0.0074 (0.0009)
Gross residential density (HU/acre) on unprotected land	-0.0672 (0.0007)	-0.0141 (0.0004)	-0.0396 (0.0004)	-0.0144 (0.0011)	-0.0391 (0.0006)
Intersection density in terms of multi-modal intersections having three legs per square mile	-0.0143 (0.0002)	0.0040 (0.0001)	-0.0078 (0.0001)	-0.0112 (0.0003)	-0.0010 (0.0001)
Number of jobs per household	0.0158 (0.0002)	-0.0040 (0.0001)	-0.0052 (0.0001)	-0.0156 (0.0003)	-0.0157 (0.0002)
Regional Diversity*	0.1822 (0.0046)	-0.2039 (0.0026)	-0.1312 (0.0024)	-0.1877 (0.0076)	-0.1952 (0.0041)

* regional diversity measures the deviation of the CBG employment rate (jobs per person) from the regional average employment rate.

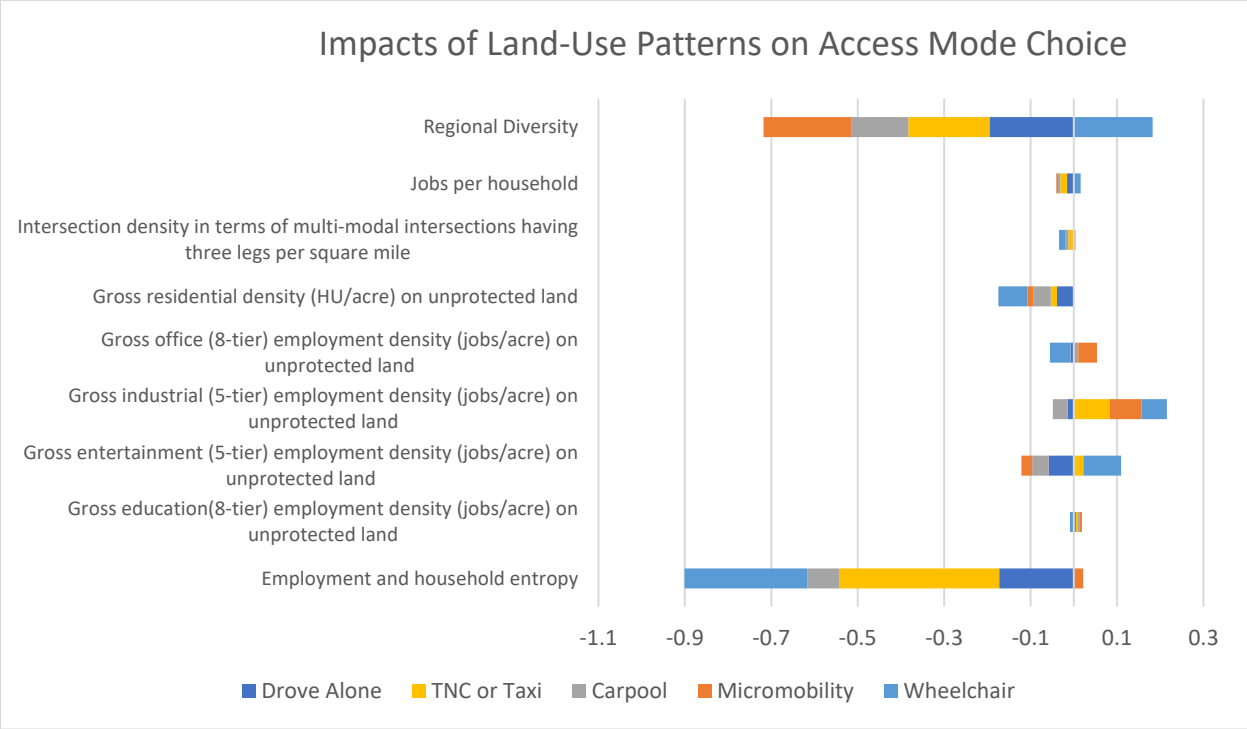


Figure 19 Relative Impacts of Land-Use Patterns

1.4.2.2 Egress Mode Choice Model

Impacts of Trip Characteristics

Table 9 below presents the model results of the egress mode choice model in terms of trip attribute variables. It shows that as egress length increased, the probability of using motorized mode and micromobility modes increased as well. In terms of the impacts of destination place type, users going to medical visits or hospitals were more likely to drive from transit stops to their destinations, while college/university students were less likely to carpool. Users going to shopping places were less likely to use micromobility and carpool modes for last-mile connection. On the other hand, users coming from airports were more likely to use carpool (including being picked up) and TNC or taxi modes for egress trips, and those coming from medical visits or hospitals showed a higher propensity of using the wheelchair or drive alone mode for the egress link.

Similar to the access mode choice model, the number of transfers showed negative impacts on the probability of choosing the motorized modes for egress purpose compared to walking. Carpool and drive alone were less likely to be used in the midday period.

Table 9 Model Results for Trip Characteristics

Base Category	Parameter	Wheelchair	Micro-mobility	Carpool	TNC or Taxi	Drove Alone
	Intercept	-13.4066 (7.9537)	-4.7544 (0.3286)	-3.6420 (0.3349)	-6.7790 (0.7589)	-5.1356 (0.6185)
Egress Length	**Distance from a transit stop to the destination (mile)**		1.3488 (0.0785)	1.9382 (0.0730)	1.9516 (0.0743)	1.9496 (0.0738)
Destination Place	**Medical, Hospital**					1.6571 (0.5022)
	Shopping		-0.5440 (0.2947)	-0.8199 (0.4143)		
	University/college			-8.2498 (1.5404)		
Origin Place	**Airport**			1.6083 (0.3247)	1.2316 (0.6448)	
	Medical, Hospital	1.3308 (0.5081)				1.2256 (0.5121)
Transfer	**Number of transfers from the origin**		-0.2695 (0.1084)	-0.3085 (0.1304)		-1.7543 (0.3952)
	Number of transfers to the destination		-0.3313 (0.1084)	-0.3416 (0.1226)	-2.0377 (0.6283)	-1.3293 (0.3030)
Two-way trip	**Trip in the Opposite Direction-Yes**				-0.6058 (0.3155)	
Time Period	**Midday**			-0.5634 (0.1478)		-1.2591 (0.3002)
Month	**February**			-0.3214 (0.1499)	-0.8936 (0.4040)	

Figure 20 presents the impacts of trip characteristics on egress mode choice. Egress length showed great impacts on the use of all four motorized modes. College/university trips showed the largest negative impact on carpooling. The number of transfers showed great negative impacts on the use of TNC or taxi or drive alone modes. Medical or hospital trips (either as the origin or destination) were more likely to use drive alone mode for last-mile connection.

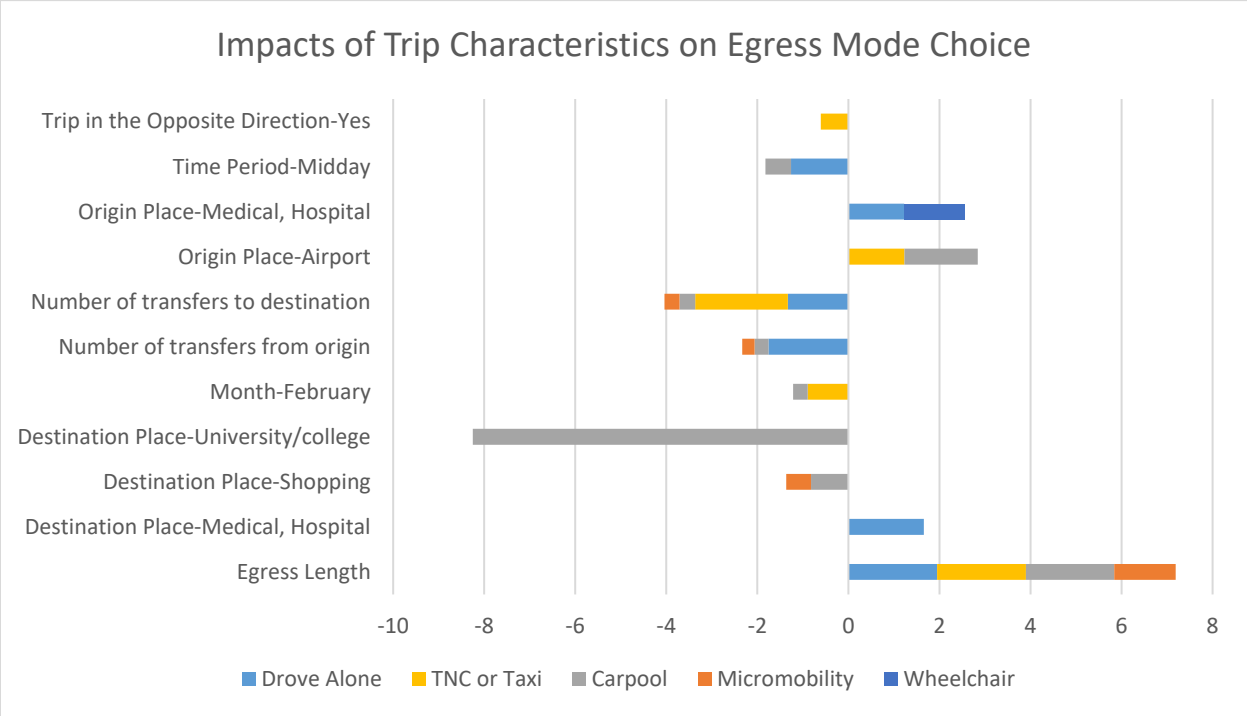


Figure 20 Relative Impacts of Trip Characteristics

Impacts of User Characteristics

Table 10 presents the model results for user characteristics in terms of their impacts on egress mode choice. Users with disabilities showed a higher probability of carpool or being picked up by others than using other modes. Those with driver's licenses were more likely to drive or use TNC or taxi for their egress trips.

Looking at ethnicity, African Americans and Hispanics were more likely to walk; Asians showed a higher propensity of using TNC or taxi services, while American Indians showed higher preferences of micromobility and less preference or carpooling for egress linkage. Interestingly, male users showed a lower probability of using drive alone mode than female users.

In terms of household income, middle-income ($20K-$50K) users were more likely to choose micromobility, followed by TNC or taxi. Users with high household income showed a high inclination to use TNC or taxi services for the last mile, followed up drive alone, carpool and micromobility.

Figure 21 presents the impacts of user characteristics.

Table 10 Model Results for User Characteristics

Base Category	Parameter	Wheelchair	Micro-mobility	Carpool	TNC or Taxi	Drove Alone
	Intercept	-13.4066 (7.9537)	-4.7544 (0.3286)	-3.6420 (0.3349)	-6.7790 (0.7589)	-5.1356 (0.6185)
Disability	**Disability-Yes**			0.3710 (0.2085)		
Driver License	**Driver License-Yes**				0.9973 (0.3394)	2.7107 (0.4122)
Ethnicity	**African American**		-0.4305 (0.1279)	-0.5889 (0.1513)	-0.8613 (0.3531)	-1.7695 (0.2962)
	American Indian		0.8178 (0.4512)	-1.7185 (1.0373)		
	Asian				1.4318 (0.5830)	
	Hispanic		-0.3602 (0.1457)	-0.2862 (0.1597)		-1.1175 (0.2841)
Gender	**Male**					-0.4788 (0.2099)
HH Income	**Middle Income ($20K-$50K)**		0.5542 (0.1197)		0.9792 (0.3493)	
	High Income ($50K-$100K)		0.3693 (0.2035)	0.7323 (0.1930)	1.4262 (0.4377)	0.9137 (0.2805)

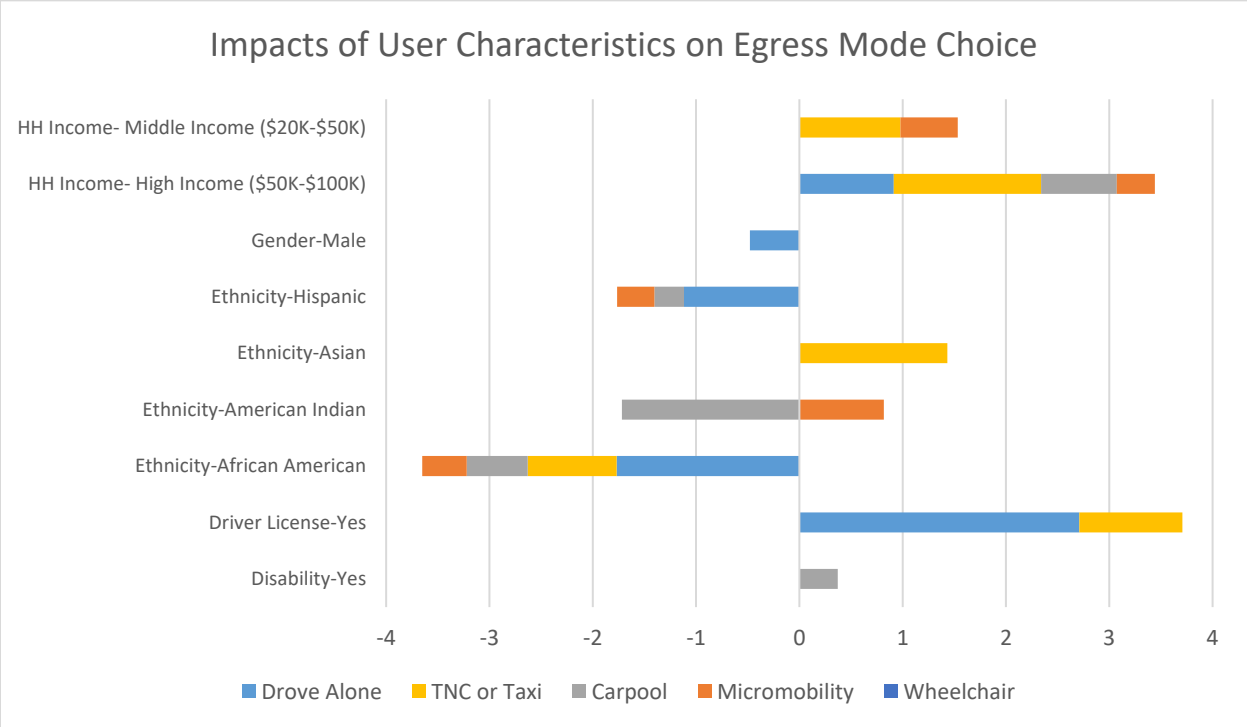


Figure 21 Relative Impacts of User Characteristics

Impacts of Land Use Characteristics

Table 11 presents the model results for land use patterns at the destination in terms of the impacts on the choice of egress mode. It shows that private car related modes (carpool or drive alone) were less likely to be used at destinations with more households with zero auto ownership. Entertainment employment density also showed negative impacts on the use of drive alone or carpool modes. Destinations with higher proportional accessibility also showed strong positive impacts on the use of micromobility modes, also drive alone and carpool use. Destinations with high regional diversity also encouraged the use of TNC or taxi services for egress purposes. Figure 22 presents the impacts of the land-use variables. Accessibility and diversity measures showed strong impacts on the egress mode choice behavior.

Table 11 Model Results for Land-Use Patterns

Parameter	Wheelchair	Micro-mobility	Carpool	TNC or Taxi	Drove Alone
Intercept	-13.4066 (7.9537)	-4.7544 (0.3286)	-3.6420 (0.3349)	-6.7790 (0.7589)	-5.1356 (0.6185)
Number of households in Destination CBG that own zero automobiles			-0.0016 (0.0007)		-0.0042 (0.0014)
Gross entertainment (5-tier) employment density (jobs/acre) on unprotected land of destination			-0.0880 (0.0427)		-0.2173 (0.1125)
Proportional Accessibility to Regional Destinations		3.9741 (1.3076)	3.8362 (1.1854)		4.8458 (1.2974)
Regional Diversity of destination				1.4658 (0.4872)	0.7102 (0.3621)

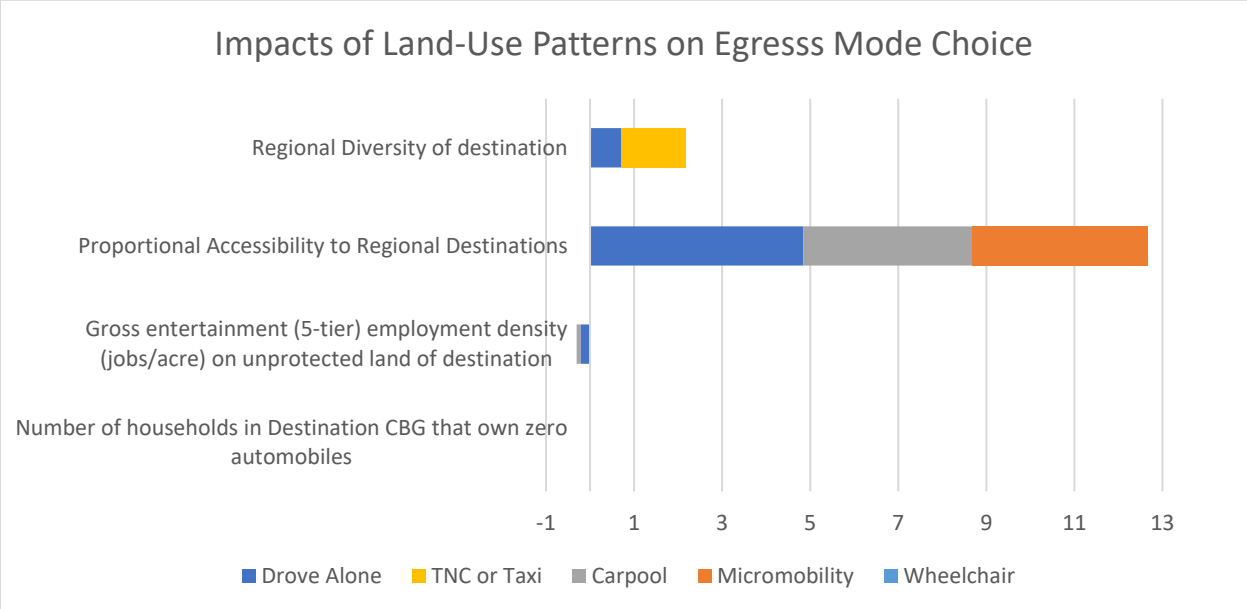


Figure 22 Relative Impacts of Land-Use Patterns

1.5 CONCLUSION

This report presents a study aiming to investigate the potential market of TNCs to serve as the first/last mile connection for transit services. To achieve this goal, this study investigated the influential factors that affect transit users' choices of access and egress modes, including TNC or taxi, drive alone (PNR), carpool (KNR, carpool or shuttle), micromobility modes (bike-sharing, scooters, etc.), and others. Transit on-board survey data collected in Spring 2017 for the Orlando metropolitan area were used for this analysis. The survey covered bus services provided by LYNX. The survey collected detailed trip information for all segments of the transit trips, including access and egress links. User demographics and household information were also recorded in the survey. This data provided the opportunity to look into the characteristics of transit trips and investigate the influential factors to users' mode choice behavior for first and last-mile connections.

In addition to the transit on-board survey data, various other data were also compiled and integrated into the GIS database to facilitate the analysis. These data include GTFS data for LYNX transit networks, street data, and the Smart Location Database (SLD) provided by EPA. The SLD data provided a comprehensive set of land use attributes at census block group level, including population and employment information, density measures, diversity variables, land use design variables, transit-related attributes, and accessibility measures. These data provided the opportunity to investigate how land use characteristics may contribute to users' choice for access and egress modes, beyond the personal and household attributes.

Separate multinomial logit models (MNL) were developed to investigate the mode choice for access and egress links, respectively. The models revealed interesting insights into transit users' choice behavior for first and last-mile connections. Various personal characteristics, trip attributes, and land use variables showed significant impacts. Particularly, trips going to airports or universities/colleges had much higher probabilities of using TNC for access and egress purposes. On the other hand, sports events and medical visits were less likely to be connected through TNC services. Visitors and evening trips were also less likely to start with TNCs. A longer distance between the origin and the transit service showed positive impacts on the use of TNC services. Higher household income also showed a positive influence on TNC usage.

In view of land use characteristics, higher employment and household entropy and higher diversity at the origin showed positive impacts on the use of micromobility and walking, and reduced the probability of using motorized modes, including TNCs, for access purpose. On the destination side, higher diversity seemed to encourage the use of TNCs and drive alone modes for egress purposes.

This study provides useful insights into the factors that may influence transit users' choice of modes for access and egress purposes. It may help transit agencies and planners in

understanding the potential market for using TNCs as first/last mile connections for transit services.

1.6 REFERENCE LIST

1. Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), 716-723.
2. Alemi, F., & Rodier, C. (2018). Simulation of Ridesourcing Using Agent-Based Demand and Supply Models Regional: Potential Market Demand for First Mile Transit Travel and Reduction in Vehicle Miles Traveled in the San Francisco Bay Area (No. 18-05283).
3. Ben-Akiva, M. E., Lerman, S. R., & Lerman, S. R. (1985). Discrete choice analysis: theory and application to travel demand (Vol. 9). MIT press.
4. Berrada, J., Andreasson, I., Burghout, W., & Leurent, F. (2019). Demand modeling of autonomous shared taxis mixed with scheduled transit. In 98th Annual Meeting of the Transportation Research Board.
5. Curtis, T., Merritt, M., Chen, C., Perlmutter, D., Berez, D., & Ellis, B. (2019). Partnerships Between Transit Agencies and Transportation Network Companies (TNCs) (No. Project J-11/Task 26).
6. Davidson, A., Peters, J., & Brakewood, C. (2017). Interactive travel modes: Uber, transit, and mobility in New York City (No. 17-04032).
7. EPA (2014) Smart Location Database Version 2.0 User's Guide. Retrieved from: https://www.epa.gov/smartgrowth/smart-location-database-technical-documentation-and-user-guide
8. EPA (2020) Smart Location Mapping. Retrieved from: https://www.epa.gov/smartgrowth/smart-location-mapping
9. Farhan, J., Chen, T. D., & Zhang, Z. (2018). Leveraging shared autonomous electric vehicles for first/last mile mobility (No. 18-06633). systems with public transit (No. 17-05439).
10. Greene, W. H. (2003). Econometric analysis. Pearson Education India.
11. GTFS (2020) GTFS: Making Public Transit Data Universally Accessible. Retrieved from: http://gtfs.org/
12. Hall, J. D., Palsson, C., & Price, J. (2018). Is Uber a substitute or complement for public transit?. Journal of Urban Economics, 108, 36-50.
13. Jaller, M., Pourrahmani, E., Rodier, C. J., & Bischoff, J. (2019). Simulation-Optimization Framework to Evaluate a Sustainable First Mile Transit Access Program Using Shared Mobility (No. 19-03368).
14. Koppelman, F. S., & Sethi, V. (2000). Closed-form discrete-choice models. Handbook of transport modelling, 1, 211-222
15. Lee, D., Derrible, S., & Pereira, F. C. (2018). Comparison of four types of artificial neural network and a multinomial logit model for travel mode choice modeling. Transportation Research Record, 2672(49), 101-112.

16. LYNX (2020a) GIS/GTFS Data Download. Retrieved from: https://www.golynx.com/maps-schedules/data-download.stml
17. LYNX (2020b) NeighborLink. Retrieved from: https://www.golynx.com/plan-trip/riding-lynx/neighborlink.stml
18. McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.
19. Mohammed, E. A., Naugler, C., & Far, B. H. (2015). Emerging business intelligence framework for a clinical laboratory through big data analytics. Emerging trends in computational biology, bioinformatics, and systems biology: algorithms and software tools. New York: Elsevier/Morgan Kaufmann, 577-602.
20. Pinto, H. K. R. D. F., Hyland, M. F., Verbas, İ. Ö., & Mahmassani, H. S. (2018). Integrated mode choice and dynamic traveler assignment-simulation framework to assess the impact of a suburban first-mile shared autonomous vehicle fleet service on transit demand (No. 18-06707).
21. Schwanen, T., & Mokhtarian, P. L. (2005). What affects commute mode choice: neighborhood physical structure or preferences toward neighborhoods? Journal of transport geography, 13(1), 83-99.
22. Shen, Y., Zhang, H., & Zhao, J. (2017). Embedding autonomous vehicle sharing in public transit system: An example of last-mile problem (No. 17-04041).
23. Stiglic, M., Agatz, N., Savelsbergh, M., & Gradisar, M. (2018). Enhancing urban mobility: Integrating ride-sharing and public transit. Computers & Operations Research, 90, 12-21.
24. Stone, M. (1979). Comments on model selection criteria of Akaike and Schwarz. Journal of the Royal Statistical Society. Series B (Methodological), 276-278.
25. SunRail (2020) GTFS Data Download. Retrieved from: https://corporate.sunrail.com/about-sunrail/gtfs-data-download/
26. US Census Bureau (2020) TIGER/Line Shapefile. Retrieved from: https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html
27. Vakayil, A., Gruel, W., & Samaranayake, S. (2017). Integrating shared-vehicle mobility-on-demand
28. Washington, S. P., Karlaftis, M. G., & Mannering, F. (2010). Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC.
29. Yan, X., Levine, J., & Zhao, X. (2019). Integrating ridesourcing services with public transit: An evaluation of traveler responses combining revealed and stated preference data. Transportation Research Part C: Emerging Technologies, 105, 683-696.

2.0 TASK 2: ANALYSIS FOR SUPPLY MARKET

2.1 INTRODUCTION

In recent years, ridesharing services, such as Uber, Lyft, Avego (Carma), SideCar, DiDi (in China), etc., has quickly spread in popularity (Xu et al., 2015) in different cities and countries. It has an immediate impact on public transit. Some studies show that shared-mobility service is a complement to public transit as it promotes the ridership of transits and helps reduce traffic congestion (Feigon and Murphy, 2016). Other studies demonstrated that ridesharing mode is undermining public transportation and becoming a major contributor to snarled traffic congestion and carbon emissions (Hill, 2018). These mixed findings indicate that ridesharing and public transit present complicated competitive/complementary relationships over a spatiotemporal space and none of them can fully satisfy the modality requests individually. Forming hybrid urban public transport services such as hybrid transit system (Koffman, 2004) or cooperating with transit and emerging on-demand services (i.e., microtransit or ridesharing services) (Boarnet et al., 2017) is well accepted, both of which seek to inject or integrate the flexibility into the public transit system. Over decades, many models (Aldaihani et al., 2004; Fu, 2002; Quadrifoglio and Li, 2009) have been developed to implement and operate hybrid transit systems in different ways, but few succeeded. The main challenge is the traffic demand variation in different levels (Velaga et al., 2012). To effectively operate a hybrid system, we need to not only penetrate the gradual evolvement of transit demand to properly refine the fixed transit routes but also predict the ad hoc demand online to timely implement flexible services. In addition, the prediction and accommodation of these two types of demands should be coordinated rather than be independently conducted.

However, the majority of current studies usually predict the demand for one of them individually. For example, using the transit ridership data, demographic survey combined with land use, many studies estimated transit demand for planning the inflexible transit routes in long term (Boyle, 2006; Hashemian, 2002; Huang, 1996; Nazem et al., 2011; Roberts, 1985; Sung et al., 2014). These traditional data and analysis approaches are either too expensive or not sensitive enough to capture gradual and mild transit demand changes over weeks or months resulting from many factors, such as people's activity and seasonal requirements. Accordingly, current transit routes usually will not change over months or years, and they do not coordinate well with the flexible routes provided by on-demand services. Other efforts predicted the dynamic ad hoc demand through ridesharing data (Faghih et al., 2019; Liu et al., 2019; Xu et al., 2017; Zhang et al., 2019), i.e., Uber and taxi, only for improving ridesharing service rather than transit routes. Therefore, the state-of-the-art indicates that transit and ridesharing service data are often individually analyzed with separate objectives for the respective modes. Their prediction on demand provides limited help to coordinate the flexible and inflexible routes in a hybrid urban public transport service system. From another point of view, public transit and ridesharing services together make up a hybrid system. The demand and service gaps of such hybrid system can be sensed by their service data. Therefore, in order

to well coordinate with their services, we should put transit and ridesharing service data together for the spatiotemporal demand analysis.

In view of the above issues, this study seeks to investigate the data analysis approaches, which combine unite transit and ridesharing service data together to explore the service gaps for setting new transit stations and routes as well as integrating ridesharing and microtransit routes. The objective is to promote the cooperation between public transit and on-demand services. To do that, this study postulates that current competitive/complementary relationships between traditional transit system and ridesharing service make ridesharing vehicles function as the probes to detect the deficiency of the flexibility and coverage in the existing transit network and its services. The findings can provide clues to refine the backbone transit network and promote cooperation between transit and on-demand services.

A majority of the existing studies used the ridership data (including the pickup and drop-off data of ridesharing service or loading and alighting data of transit service) for the demand prediction. Few studies investigated the trip data collected from both ridesharing and transit sides, which may reflect the deficiency of the flexibility and coverage in existing transit routes from a different perspective. For example, the spatiotemporal areas dominated by ridesharing trips (services) indicate potential deficient transit services, while other areas evenly covered by both transit and ridesharing trips indicate diverse traffic demand for hybrid service modes. Thus, putting the trip data from both modes together will give us a new angle to discover the potential supply-demand market for integrating these two modes in a hybrid service system. However, it is not trivial to implement the aforementioned data analysis. We introduce the research challenges along with our contributions to the solution approach development as follows.

First, the transit trips have fixed routes and schedules, while the ridesharing services randomly distribute over a city during different time slots over a day. The two sets of trip data together ramblingly scatter in a spatiotemporal space and they are non-additive, which make many quantitative approaches have no way to start directly. To address this difficulty, we first developed a new data presentation approach, which considers each trip as a 3D curve and then meshed their spatiotemporal services with an optimal 3D grid involving a number of uniform 3D cubes. This optimal 3D discretization enables us to zoom in and study the service competition between these two modes in each cube. However, the results only provide information fragments rather than the service gaps that we are interested in. This problem motivates our study to further aggregate the cube information fragments to ridesharing/transit swarms – the areas formed by the connected cubes which are dominated by one service – in each time slice. Built upon that, we are able to explore the corresponding spatial service gaps and provide future route planning suggestions. Next, by combining the ridesharing pick-up, drop-off data and transit station data, we propose an innovative approach – identifying “sandwich” patterns – to locate the potential first/last mile zones for integrating microtransit services. Last, we recognize that the insights obtained at each time slice can only present spatial demand variation but not temporal demand dynamics. Consequently, it will not help planning on-demand services that require the understanding of the temporal variation. This study then piles

the information slices in the temporal dimension as a time series data and feeds them into a deep learning network to predict future demand patterns.

The effectiveness of this data analysis approach is validated by a case study built upon the field data collected in the second ring region of Chengdu, China. Specifically, we find nine first/last mile zones. Combining with land-use data, we noticed that they are either big commercial or residential areas with high population density. Among them, some areas locate around metro lines, thus microtransit service is potentially needed to cover first/last mile demand. Some areas are away from the metro line but with low transit service and they potentially require additional transit service coverage. Thus, analyzing these trip data provides constructive guidance to improve the current transit service by considering the ridesharing services.

We summarize the main contributions of this study as follows: (i) We developed a new 3D presentation approach to present the trip data collected with different collection rates and coverage; (ii) We developed an innovative data analysis approach, which spatially aggregates information fragments and temporally piles the spatial information to uncover the potential transit service gaps hidden in trip data involving both transit and ridesharing modes; (iii) We found two interesting patterns, service swarms and "sandwich" patterns, which respectively point out underlying transit demand and first/last mile zones suitable for potential microtransit service; (iv) Using deep learning method, we predicted the time-vary transit first/last mile zones which helps the operation of microtransit service; and (v) We analyzed a set of field data collected from Chengdu city in China and validated the effectiveness of our approaches. Our analysis of ridesharing trip data in this study can be extended to other mobility modes, such as bike-sharing and private vehicle trip data, to provide a thorough understanding of transit service gaps. These contributions together benefit the development and operations of hybrid urban public transport systems.

2.1.1 Objective

This task assesses the spatiotemporal service gaps of transit services. The objective is to investigate when and where are the supply gap/hubs to either integrate shared mobility and public transit services or properly implement hybrid transit systems.

2.1.2 Scope

This task focuses on data analysis of transit trip data and ridesharing trajectory data in the second ring region of Chengdu, China in 2016.

2.2 LITERATURE REVIEW

This study focuses on developing innovative approaches to analyze trip data collected from both transit and ridesharing services, aiming to discover the mobility service gaps and their variation in an urban area. The findings of this study will help promote the cooperation of transit system and existing on-demand services for improving the service level in an urban public transport system. This research is closely related to the demand prediction for transit

and ridesharing. Our review will introduce these closely related studies in literature, differentiate this effort from existing studies and further highlight our contributions.

We first discuss the state-of-the-art of transit demand prediction. According to Boyle (2006), existing studies mainly applied ridership (Fang et al., 2018; Hashemian, 2002; Huang, 1996; Jun et al., 2015; Nazem et al., 2011; Nourbakhsh and Ouyang, 2012; Roberts, 1985; Sung et al., 2014) and O-D survey data (Chatterjee and Venigalla, 2004) to predict transit demand. For example, (Nazem et al., 2011) analyzed the travel patterns of different demographic classes to understand the relationship between transit ridership and demographics. (Sung et al., 2014) employed spatial regression analysis to investigate the impact of land use on the rail transit ridership in the city of Seoul. (Jun et al., 2015) applied a multinomial logit model to analyze how land use and demographic characteristics affect transit ridership. In recent years, the ridership data collected by the Automated Fare Collection (AFC) system is used to capture the variation of the transit demand, especially for railway system (Fang et al., 2018). For example, based on AFC data, (Nourbakhsh and Ouyang, 2012) developed the state-space model to predict the real-time subway demand, considering the impact of special events. In the meantime, extensive studies analyzed the ridership data collected from ridesharing services, but mainly for predicting the ridesharing demand. For example, based on Uber pick-up data, (Faghih et al., 2019) applied the LASSO spatial-temporal autoregressive model to predict the Uber demand in Manhattan. (Xu et al., 2017) fed the taxi pick-up and drop-off data in New York City into a long short-term memory (LSTM) neural network to forecast the future taxi requests. (Zhou et al., 2018) employed the convolutional LSTM (ConvLSTM) to capture the spatiotemporal relationship of both taxi and bike-sharing demand data in New York for a short-term demand prediction. (Zhang et al., 2019) developed an end-to-end multi-task learning temporal convolutional neural network to predict the short-term ridesharing demand and compared its performance with the state-of-the-art deep learning approaches.

This brief review indicates several research gaps that this study tries to make up. First of all, the majority of existing studies predicted/estimated the passenger demand for transit or ridesharing services through their own ridership data (i.e., transit smart card data or ridesharing pick/drop data) combined with the demographic and land-use features. However, few studies, like this research, investigated the service trip data collected from both transit and ridesharing. This study noticed that these trips can reflect the collective and dynamic competition between these two modes in a spatiotemporal space. Consequently, analyzing their trip data may offer unique insights for integrating these two types of mobility services in a local network. To the best of our knowledge, this is the first attempt to investigate the service deficiencies of an existing transit system through analyzing its competitors' services such as ridesharing trip data. On the other hand, the trip data are non-additive curves spanning in a local network during a period. Existing approaches, such as various choice models and regression analysis, which have been successfully used to analyze ridership, land use and demographic data, cannot be directly applied to study trip data. It calls for new approaches to conduct the data analysis. The above points highlight the novelty and unique methodology contribution of this study.

From the application view, this study will significantly contribute to the development of the hybrid transit system or the integration of transit system with emerging on-demand services. Even though the concept of the hybrid system has been proposed for decades, only a small percentage of transit agencies (Potts et al., 2010) adopted it. The uncertainty of passenger demand (Velaga et al., 2012) plays one of the critical challenges. For example, (Qiu et al., 2014) indicated that the uncertainty of demand leads to inappropriate slack time and makes flex-route service fragile and inefficient. Therefore, to operate a hybrid transit system well, we need new research approaches to better understand the demand varying in different levels in a hybrid mobility service environment. Few studies in the literature investigate this research need. This study seeks to partially fill in this research gap, too. Below is the unique methodology developed by our team of researchers, which we believe contributes significantly to the state-of-the-art.

2.3 METHODOLOGY

This study is devoted to the development of innovative approaches to analyze trip data collected from transit and ridesharing services to investigate the flexibility and coverage gaps of the current transit system. These findings will help implement new fixed routes, stations as well as microtransit to promote the implementation of a hybrid transit system. To do that, we consider ridesharing trips as the detectors to reveal the potential mobility demand for the flexible and inflexible transit routes and also explore their evolvements over different temporal and spatial horizons. Along with the above thought, this study mainly considers the trip data defined as follows. We consider V number of ridesharing vehicles and B number of bus/metro services. Their trajectories are respectively updated at discrete time stamps $n \in \{0,1,\dots,N\}$ according to ridesharing vehicles GPS updating frequency, and at $m \in \{0,1,\dots,M\}$ corresponding to the bus arrival time at the stations. Accordingly, the trip of the ridesharing vehicle v is denoted as $\mathcal{Z}_v^S = \left\{z_{v,n}^S(x,y), n \in \{0,1,\dots,N\}\right\}, \forall v \in V$, where $z_{v,n}^S(x,y)$, abbreviated as $z_{v,t_n}(x,y)$, is the coordinates of vehicle v at the nth time stamp (time t_n); similarly, the trip of bus b is denoted as $\mathcal{Z}_b^T = \left\{z_{b,m}^T(x,y), m \in \{0,1,\dots,M\}\right\}, \forall b \in U$, where $z_{b,m}^T(x,y)$ is the coordinates of bus b at the time mth time stamp. We mark t_0 as the departure time at start station and $t_m, m \in \{1,\dots,M\}$ is the arrival time at following stations along the transit line.

Built upon the above data, this study will develop data analysis approaches, which integrate spatiotemporal statistical data analysis, machine learning and optimization approaches, to provide the following capabilities. We will first discover when and where transit and ridesharing compete or complement on the local transportation network. Based on this knowledge, we will reveal transit mobility service deficiencies in flexibility and coverage as well as their temporal and spatial variation pattern so that we can provide valuable planning suggestions to operate a hybrid transit system. Our data analysis approaches include three key components: developing optimal discrete 3D presentation, analyzing ridesharing service swarms to discover spatial transit service gaps and then predicting the dynamic patterns of the service gaps.

2.3.1 3D presentation and Optimal Discretization

As we mentioned above, this study involves ridesharing GPS trajectory data and bus trip data. These two sets of data ramblingly scatter in the traffic network and their coverage change over time, i.e., each trip starts and ends at different time and locations by going through different roads. To uncover the service pattern involved in the data, we need a good presentation to support the analysis. Considering the spatiotemporal dynamics of the trip data, this study puts all trips in a 3D space spanned by 2D (x-y) spatial coordinates and time (t) dimension. This 3D space represents the entire service space. Accordingly, each individual trip is presented by a 3D route in the space. See Figure 23 for an example.

Furthermore, we noticed that the 3D routes intersect and then diverge at different spatiotemporal points. Some areas present very dense trips going through by both bus and ridesharing modes, where they compete for demand, but others sparsely visited by one of them, where they complement. Most importantly, these trips data are not directly addable. To systematically analyze these competitive and complementary relationships between these two traffic service modes, this study discretizes the 3D service space into $K \times I$ number of uniform cubes (see Figure 23, where K is the number of the pixels spatial area and I is the number of time interval in dimension. And then, we examine their service relationship in each cube first by observing the number of bus or/and ridesharing services occurring in each cube. Built upon the analysis in each cube, we propose statistical and machine learning methods to discover the complementary or competitive pattern over the entire 3D space. To do that, it is noticed that the determination of cube dimension (i.e., time interval and the pixel size in spatial dimension) is therefore very critical to the analysis. Different cube sizes may lead individual cubes to present different information, which provides a different interpretation for the competition relationship. This study next discusses the significance of this factor in detail and then presents our approach to decide the optimal length of the time interval and pixel size.

2.3.1.1 Optimal Time Interval

The data analysis based upon the 3D presentation first discretizes (slices) the study horizon by a fixed time interval τ. This section investigates how the length of the time interval of each cube will affect the statistical analysis and then explores the optimal time interval τ^*. Please note that with a given time interval, we locate the position of a ridesharing vehicle by its averaged coordinates during each time interval. This process will compromise the accuracy of the location information. It also may lead to miscounting of the number of ridesharing services in each cube if the cube size is not proper. For example, the actual trajectory of a ridesharing vehicle may go across multiple cubes during an interval, and then this ridesharing service should be counted in each of these cubes. However, by taking the average coordinates, it can only fall in one cube but not other cubes. If the time interval is too wide, it may cause significant miscounting. On the other hand, if the time interval is too small, it will lead to a large number of cubes and makes the training process expensive. Therefore, the selection of τ needs to balance the computation load and the information accuracy. To address this dilemma, we develop the optimization model in (4) -(7), which explores the optimal time interval with the objective to minimize the information loss and the dataset size, subject to feasible range of the

value τ. Note that transit service data is not involved in this optimization model, because we analyze the arrival time of transit vehicle at each stop rather than transit trajectory, the accuracy of transit service location information is not influenced by improper time discretizing.

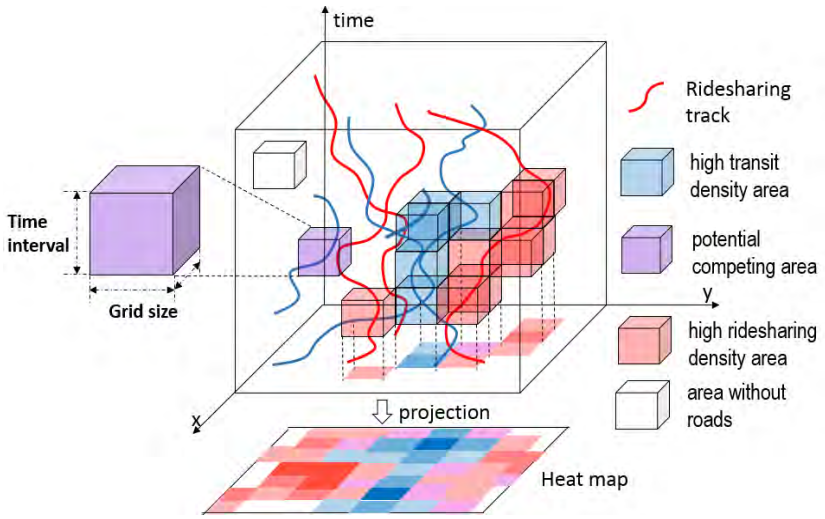


Figure 23 3D presentation

More exactly, this study uses the sample variances (the first item in (4)) to measure the loss of the location information and also puts the penalty on the number of the cubes. Note that this optimization model is hard to be solved because of the presence of the decision variable τ underneath the summation symbol. However, it only involves one variable. In practice, we can find a lower and upper bound for τ so that the searching space is reasonably limited. For example, in our dataset, the DiDi vehicle trajectory data has updating frequency of 2~4 seconds which makes $\underline{\tau} = 4$; the average DiDi single trip service time provides an upper bound $\bar{\tau}$. Therefore, we can quickly search the local optimal solution of τ by using line search approaches (Dechter and Pearl, 1985) in a narrow feasible region $[\underline{\tau}, \bar{\tau}]$.

$\boldsymbol{P_1}$:

$$\min_{\tau} \alpha \sum_{i=0}^{I} \frac{1}{|U^i|} \sum_{v \in U^i} \frac{1}{|\mathbb{Z}_{v,i}^{\tau}|} \sum_{i\tau \le t \le (i+1)\tau} ||\boldsymbol{z}_{v,t}(x,y) - \overline{\boldsymbol{z}_{v,t}}(x,y)||_2^2 + \beta M \frac{T}{\tau} \tag{4}$$

Subject to:

$$\underline{\tau} \le \tau \le \bar{\tau}, \tag{5}$$

$$I = \left\lfloor \frac{T}{\tau} \right\rfloor, \tag{6}$$

$$\overline{\boldsymbol{z}_{v,t}}(x,y) = \frac{1}{|\mathbb{Z}_{v,i}^{\tau}|} \sum_{i\tau \le t \le (i+1)\tau} \boldsymbol{z}_{v,t}(x,y), \forall v \in V_t, t \in T, \tag{7}$$

where τ, the length of the time interval, is the single decision variable and $\boldsymbol{z}_{\boldsymbol{v},\boldsymbol{t}}(x,y)$ is the input data, which represents the location coordinates of the v-th vehicle at time t; $i \in I$ is the index for time interval; V_t represents the set of ridesharing vehicles at time t. U^i represents the set of vehicles during i-th interval; $\mathbb{Z}_{v,i}^{\tau} \subset \mathcal{Z}_v^S$ is the set of coordinate records of v^{th} ridesharing vehicle during the time interval $[i\tau, (i+1)\tau]$; specifically, $\mathbb{Z}_{v,i}^{\tau} = \{\boldsymbol{z}_{\boldsymbol{v},\boldsymbol{t}}(x,y) | i\tau \le t \le (i+1)\tau\}$; α and β are the predefined weights normalized to make the two terms comparable in the objective function in the magnitude. An optimal solution of τ enables the data analysis proposed in this study to hold both desired information accuracy and acceptable computation load.

2.3.1.2 Optimal Pixel Size

This study next determines the pixel size of the cubes in the spatial dimension. In addition to the time interval τ, the size of the pixel will also affect the number of the cubes as well as the counting of the ridesharing/transit services occurring in each cube. Thus, it will influence the computation load as well as interpretation power of the data analysis. See Figure 24 for an example. We consider the average speed of the traffic on a road is v_a and the speed upper limit is $\overline{v}$. With the given time interval width τ, if the size of the pixel is too small, such as the length of the edge $l < \tau v_a$ in Figure 24 (a), it will very likely lead to many empty pixels since the majority of vehicles can run across a pixel during a time interval τ. On the other hand, if the size of the pixel is too large, such as the length of the edge $l > \tau\overline{v}$ in Figure 24 (b), it will lead to overcount since an individual vehicle will have more than one record in the pixel.

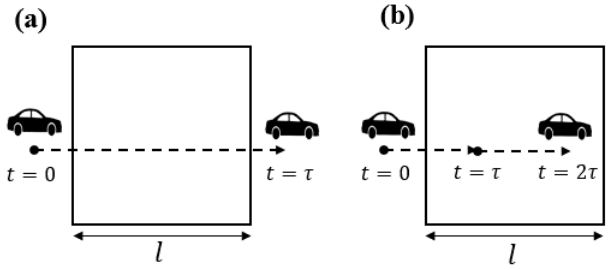


Figure 24 Examples of improper pixel size. (a) $l < \tau v_a$, (b) when $l >$

The example above shows that a pixel with the length within $[v_a\tau, \overline{v}\tau]$ will facility the analysis better. An improper discretization will lead to many empty cubes or overcount. It will notprovide valuable insights and will affect the statistical analysis significantly. Therefore, we seek to explore the optimal pixel size so that they can clearly present either complementary or competitive relationships between traffic modes. To do that we first formally define a complementary or competitive pixel in Definition 1 and 2 through the ridesharing service ratio defined in Equation (8) below.

$$R = \{r_{k,i}\}, r_{k,i} = \begin{cases} \frac{N_{k,i}^S}{N_{k,i}} & N_{k,i} > 0 \\ -1 & N_{k,i} = 0 \end{cases}, k \in K, i \in I, \quad (8)$$

where $N_{k,i}^{s}$ and $N_{k,i}$ respectively denote the number of ridesharing trips and total trips going through pixel k during the i-th time interval. Equation (8) indicates that $r_{k,i} \in [0,1]$ for $N_{k,i} > 0$ and we mark $r_{k,i} = -1$ if it is an empty pixel (i.e., no trips present in the pixel).

Definition 1 - Complementary Pixel: A pixel presents a complementary relationship between the ridesharing and transit services, if and only if its ridesharing service ratio satisfies $r_{k,i} \in \left[0, \underline{\eta}\right] \cup [\overline{\eta}, 1], k \in K, i \in I,$ where $\underline{\eta}$ and $\overline{\eta}$ are given parameters and $\overline{\eta} = 1 - \underline{\eta}$.

Definition 2 - Competitive Pixel: A pixel presents a competitive relationship between the ridesharing and transit services if and only if its ridesharing service ratio satisfies $r_{k,i} \in \left[\underline{\mu}, \overline{\mu}\right], k \in K, i \in I$, where $\underline{\mu}, \overline{\mu}$ are given parameters, $\overline{\mu} = 1 - \underline{\mu}$.

Definition 3 - Noise Pixel: A pixel cannot present a clear competitive relationship between the ridesharing and transit services if its ridesharing service ratio satisfies $r_{k,i} \in \left[\underline{\eta}, \underline{\mu}\right] \cup [\overline{\mu}, \overline{\eta}]$ (i.e., unidentifiable pixel presenting neither complementary nor competitive relationship) or $r_{k,i} = -1$ (empty pixel), $k \in K, i \in I,$ $\overline{\eta} = 1 - \underline{\eta}$ and $\overline{\mu} = 1 - \underline{\mu}$.

According to Definitions 1 and 2, a complementary pixel indicates that either the transit or the ridesharing dominates the service going through the pixel. Potentially, a group of such complementary pixels shows one of the two services is not sufficient. This information is very valuable to facilitate our data analysis later. However, a competitive pixel shows that neither the bus nor the ridesharing presents apparent merits to the trips going through the pixel. Accordingly, we consider those pixels presenting either a complementary or competitive relationship as informative pixels in contrast to the noise pixels defined in Definition 3. The ratios of informative pixels and noise pixels are affected by the schemes of the 3D discretization. We next discuss our ideas to search for the optimal discretization scheme.

First of all, our approaches are more interested in those informative pixels. The discretization therefore seeks to generate sufficient informative pixels. Accordingly, an optimization model (9)-(12) is developed with the objective to find the optimal number of pixels (κ^*), so does the size, for maximizing the total number of the informative pixels with a given time interval τ (see the objective function).

$\boldsymbol{P_2}$**:**

$$\max_{\kappa} \sum_{i=0}^{I} \left(pA_{\kappa,i} + q\bar{A}_{\kappa,i}\right) \tag{9}$$

Subject to,

$$A_{\kappa,i} = \frac{1}{\kappa} \sum_{k=1}^{\kappa} u_{k,i}\,(\kappa), \qquad \forall i \in I \tag{10}$$

$$\bar{A}_{\kappa,i} = \frac{1}{\kappa}\sum_{k=1}^{\kappa} v_{k,i}\,(\kappa), \qquad \forall i \in I \tag{11}$$

$$\underline{\kappa} \le \kappa \le \bar{\kappa} \tag{12}$$

where κ is the decision variable, representing the total number of the pixels, each with a square shape; $u_{k,i}$ and $v_{\kappa,i}$ are auxiliary binary variables, $u_{k,i} = 1$ if $r_{k,i} \in [0, \underline{\eta}] \cup [\overline{\eta}, 1]$, and 0 otherwise; $v_{k,i} = 1$ if $r_{k,i} \in [\underline{\mu}, \overline{\mu}]$, and 0 otherwise. $A_{\kappa,i}$ and $\bar{A}_{\kappa,i}$ are the proportion of the complementary/competitive pixels over the study region; p and q are predefined weights normalized to make the proportion of the two kinds of pixels comparable. Note that for a given study area with size S, the more pixels present, the smaller the pixel size is. As a proper pixel length is within $[v_a\tau, \overline{v}\tau]$, the total number of pixels, κ, is also bounded by $[\underline{\kappa}, \overline{\kappa}]$, where $\underline{\kappa} = [S/(\overline{v}\tau)]^2$ and $\overline{\kappa} = [S/(v_a\tau)]^2$.

Next, given $\overline{\eta} = 1 - \underline{\eta}$ and $\overline{\mu} = 1 - \underline{\mu}$, we notice that two of the four parameters (e.g., $\underline{\mu}$ and $\underline{\eta}$) will significantly affect the solution of the discretization since they affect the value of $r_{k,i}$ used in Definitions 1-3 so do the values of $A_{\kappa,i}$ and $\bar{A}_{\kappa,i}$ in the optimization model. More exactly, this study names the interval $\underline{\mu} - \underline{\eta}$ as an unidentifiable interval (UI). A narrow UI loosens the criteria to certify competitive or complementary pixels according to Definitions 1 and 2. Accordingly, it tends to produce a discretization solution with a few pixels (i.e., a small value of κ^*) each with a large size, which may maximize the size/number of the informative pixels but result in a low resolution, e.g., a pixel may cover some areas not presenting consistent lane-use features. On the other hand, a wide UI tightens the criteria and leads to a discretization solution with plenty of pixels (i.e., a large value of κ^*) each with a small pixel size, which may lead to more noise pixels as the cost. The determination of the parameters is highly data orientated. Thus, this study will perform a sensitivity analysis for the UI in our case study. Combining with the land-use analysis, we suggest proper values for the parameters $\underline{\eta}, \overline{\eta}, \underline{\mu}$ and $\overline{\mu}$ in the case study.

The optimization model $\boldsymbol{P_2}$ is nonlinear and nonconvex, but with a single integer decision variable κ within $[\underline{\kappa}, \overline{\kappa}]$. This study thus explores the optimal solution by heuristic approaches such as the best first search (BFS) algorithm (Dechter and Pearl, 1985), which is one of the efficient sequential search algorithms in discrete optimization. For completeness, we present the main idea of this algorithm as follows. The BFS maintains a list named OPEN, which is placed with nodes (possible solutions) to be expanded. Initially, the list of OPEN includes a set of integer solutions within $[\underline{\kappa}, \overline{\kappa}]$. Then, the solutions are evaluated through the objective function (9). The worst solution is removed from the list and the best solution is expanded to include its neighbors in the OPEN list as successors. The heuristic evaluation process is repeated until no more successors are found. The best solution remains in the OPEN list is the optimal solution.

2.3.2 Searching Ridesharing Swarm

The 3D discretization enables us to capture the competitive/complementary relationship between two modes within each individual cube. We can present this result by heatmap, in which the color of each cube represents its service rate (see the example shown in Figure 25). However, they only provide us the information debris rather than insightful information for the service gaps. This motivates us to aggregate the information that individual cubes provide to explore the converged insights. Our approach includes two steps. We first slice the 3D space by the time interval τ^* and then explore the service distribution by building the set of heatmaps $\boldsymbol{h} = \{h_i, i \in I\}$, in which the heat in each pixel k of heatmap h_i is measured by its service rate r_{ik} defined in Equation (8). A heatmap h_i exhibits the competition relationships within a time interval $[i\tau^*, (i+1)\tau^*]$ over the study region. Next, we project all heatmaps to the spatial region and aggregate the heat in each corresponding pixel k by averaging the r_{ik} over all heatmap $h_i \in \boldsymbol{h}$, i.e., $R_k = \frac{1}{I}\sum_i r_{ik}$. Thus, we obtain a new aggregated heatmap H over spatial region, in which each pixel holds the heat R_k. Through the aggregated heatmap H, this study seeks to conduct two specific analyses to discover the competitive or complementary patterns over a region and understand the transit service gaps.

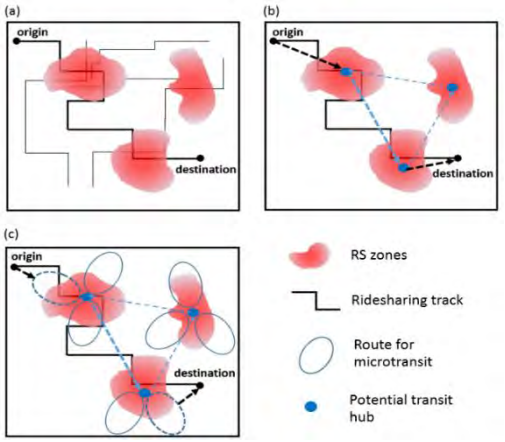


Figure 25 Searching RS zones (a) RS zones in heatmap H; (b) and (c) Potential transit hub and routes integrating ridesharing and transit services.

First, we demonstrate the capability of the heatmap H to help refine the existing transit network by improving the coverage and/or flexibility. Specifically, we mark the regions, where ridesharing services are dominant in the heatmap H, e.g., $R_k \geq 0.9$ as a ridesharing swarm (RS) zone. It means that over the aggregated period, most of the ridesharing service trips, no matter when and where their ODs are, have passed the RS zones (Figure 25 (a)). These RS regions indicate the deficiency of the coverage and/or flexibility of transit services since they attract significant ridesharing demand but have limited transit service. Consequently, if the demand is consistent, which is implied by consistent number of ridesharing trips between RS zones, we adopt the first scheme to refine the current transit backbone network which is building new stations in each RS zones and then adding new routes connecting those RS zones (Figure 25(b)). This scheme improves the network coverage, and it will potentially attract more transit

passenger demand from ridesharing. On the other hand, we also observe the case in which the demand going through an RS zone is highly dynamic. If this RS zone covers a large area surrounded by enough transit services, it indicates on-demand service requests. In this case, we propose Scheme 2: implementing microtransit services within the RS zones to accommodate the ad hoc demand around the RS zones and connect these demands to nearby transit services. Please see an illustration in Figure 25(c) where the microtransit services have flexible routes and schedules. Combining with existing ridesharing services, Scheme 2 will potentially improve transit usage. Both Schemes 1 and 2 are promising solutions to catch passenger demand around RS zones and improve the transit service. Moreover, the inclusion of these RS zones to the existing transit network generates more opportunities to improve transit ridership near/within RS zones through various modes, i.e., ridesharing, and intermodal trips (see an example in Figure 25(b), where the dashed arrow indicates an intermodal trips in which the ridesharing feeds transit service).

2.3.3 Searching First and Last Mile (FLM) gap

Suffering from the limited coverage and flexibility in the current transit system, bus passengers often meet the difficulty of the first and last mile (FLM) gap and switch to ridesharing or private auto modes. Emerging microtransit provides a promising solution to make up this deficiency by providing on-demand service for a small group of passengers. However, it is very hard to find the first and last mile gaps due to the lack of intermodal trip or relevant survey data. This study thus developed a new approach to infer the first and last mile gaps through heatmap analysis. First, given that the FLM demand usually varies from hour to hour, our analysis is built upon a

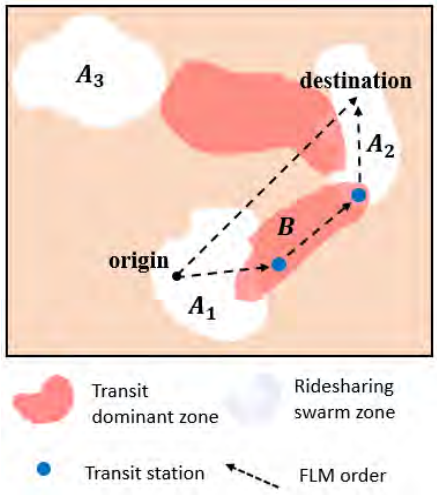


Figure 26 Schematic representation of first/last mile area pattern.

set of hourly aggregated heatmaps $\hat{\boldsymbol{h}} = \{\hat{h}_\omega, \omega \in \mathcal{W}\}$, each heatmap $\hat{h}_\omega$ is aggregated from a set of heatmaps $h_i \in \boldsymbol{h}$ over a time window of one hour. Second, we noticed that those "sandwich" patterns, in which a transit dominant zone is immediately connected by multiple ridesharing swarm zones (see Figure 26, i.e., A_1BA_2), have a great potential to present FLM zones. To interpret this thought, we show the correlation between the FLM zones and the "sandwich" patterns as follows.

First of all, we consider a large area, such as a big shopping center or residential area. If there are transit stops nearby this area, the first/last mile problem happens when the distance between the transit stops and origin/destination is beyond the walking distance. Then many ridesharing trips will show up in this area. Next, we present the correlation between the FLM zones and the "sandwich" patterns on the heatmaps. Take Figure 26 for example, the zones with white color (denoated as A zones) represent ridesharing swarm zones. The zones with red color (denoated as B zones) are transit dominant zones ($R_k < 0.5$). Both A_1 and A_2 have low transit coverage but have transit stations nearby due to their connection to B zone. In this case, the FLM problem may occur when there is a long distance between origin/destination in A_1 and/or A_2 and nearest transit station in B. Specifically, if many demand travel between A_1 to A_2 by ridersharing rather than using nearby transit line in B, this implies that the passgeners may suffer from first and last mile diffculty. Or, the demand can take intermodal trip (ridesharing and bus) to complete the trip between A_1 and A_2. In either case, A_1 and A_2 are the potential areas where ridesharing trips or microtransit can help overcome the first and last mile problem. The above analysis indicates that if we also noticed extensive ridesharing trips occuring between the zones within these patterns such as $A_1 \leftrightarrows B; A_2 \leftrightarrows B, A_1 \leftrightarrows A_2$, it is very likely that some RS zones involved in the "sandwich" patterns, i.e. A_1BA_2, are candidate FLM zones. These ridesharing orders are considered as the FLM-prone orders. The more FLM-prone orders observed, the higher possibility the RS zones within the pattern are the FLM zones.

Motivated by the above finding, we incorporate the ridesharing OD and bus station data into our analysis of the averaged heatmap $\hat{h}_\omega \in \widehat{\boldsymbol{h}}$ to search the potential FLM zones. The main idea is to first discover the "sandwich" patterns where two ridesharing swarm zones are sandwiched by a transit dominated zone on the heatmap $\hat{h}_\omega$. And then, we identify the FLM candidate zones involved in the "sandwich" patterns by referring to the ridesharing OD information. More exactly, we filter out zones that have FLM-prone orders, such as $A_1 \leftrightarrows B; A_2 \leftrightarrows B, A_1 \leftrightarrows A_2$, among "sandwich" patterns.

The zone that has more FLM-prone orders has a greater chance to be FLM zone. Therefore, for any pixel k within the RS zones of "sandwich" pattern, we consider p_k as the probability that pixel k is a FLM zone. And the probability p_k is calculated by the number of FLM orders in pixel k divided by total number of FLM orders throughout the heat map. In this way, we are able to create the time-vary FLM order probability heatmaps, $\widetilde{\boldsymbol{h}} = \{\tilde{h}_\omega, \omega \in \mathcal{W}\}$, which is the input as second channel data in the ConvLSTM model for prediction)(see the paragraph below for details). The proposed approach identifies the potential FLM zones/demand spatiotemporally, which allows transit agencies to efficiently incorporate flexibility in the services. The identified FLM zones are further validated through the bus station density analysis and land use data in the validation part of case study.

2.3.4. Learning Spatiotemporal Service Gaps

The above section manages to discover the consistent transit coverage and flexibility gaps. Relying on those observations, we provide suggestions to refine the existing transit network such as adding new bus stations, routes, or microtransit services. However, those findings are

obtained from the aggregated heatmaps along a given timeframe. They are static and provide limited help for a hybrid transit system to flexibly respond to passenger demand variation over different timeframes. To broadly incorporate the flexibility into a transit system, it is important to learn the demand dynamics. This study is thus inspired to develop the prediction model in this section.

This study uses an existing two-channel ConvLSTM learning model using the heatmap data, i.e. $h_i \in \boldsymbol{h}$ and $\tilde{h}_\omega \in \widetilde{\boldsymbol{h}}$ as inputs to predict the transit service gaps, including both ridesharing swarms and FLM zones. The following is our justification for selecting this model. One of the key characteristics of these data is the high spatiotemporal correlation. For example, some areas on heatmap h_i may share common features, such as areas near the metro hub which attract significant mobility needs including ridesharing as well as FLM demand. Moreover, the heatmaps, such as $\boldsymbol{h}$ and $\widetilde{\boldsymbol{h}}$, are sets of time series data. It's important to capture this spatiotemporal correlation in the prediction model to improve the accuracy. Recent advances in deep learning have enabled researchers to model the complex nonlinear relationships. More exactly, a convolutional neural network (CNN) has been used to capture complex spatial correlation (Zhang et al., 2016) and Long Short-Term Memory network (LSTM) has exhibited outstanding performance on time series data prediction. The ConvLSTM model, a combination of CNN and LSTM (Xingjian et al., 2015), has demonstrated the satisfied performance to capture the spatiotemporal correlation in the data for weather precipitation forecast prediction. Given the spatiotemporal characteristics of heatmap data, this study, therefore, uses the ConvLSTM model to predict the dynamics of transit service gaps.

For completeness, we briefly introduce the structure of the ConvLSTM as outlined through equation (13). The model learns sequential (temporal) correlations through a memory cell c_t. It has three gates (forget f_t, input i_t, output o_t gates) and two non-gate tanh units. The gates of ConvLSTM are able to manipulate the outputs of tanh units and value on the memory cell line, which learns temporal correlations. h_t is the output of memory cell at interval t, which is controlled by the output gate o_t. The convolution operator between weights and input data characterizes the spatial correlation.

$$\begin{aligned} i_t &= \sigma(W_i * X_t + U_i * h_{t-1} + b_i) \\ g_t &= \tanh(W_g * X_t + U_g * h_{t-1} + b_g) \\ f_t &= \sigma\big(W_f * X_t + U_f * h_{t-1} + b_f\big) \\ o_t &= \sigma(W_o * X_t + U_o * h_{t-1} + b_o) \\ c_t &= f_t \circ c_{t-1} + i_t \circ g_t \\ h_t &= \tanh(c_t) \circ o_t \end{aligned} \tag{13}$$

where b's are biases for the respective gate. W's and U's are forward and recurrent weights, respectively. They are all learnable parameters. "$\circ$" denotes element-wise multiplication, and "$*$" denotes the convolution operator.

Moreover, recall that the FLM heatmap, $\tilde{h}_\omega$, is generated upon the analysis of OD information and heatmaps $\widehat{\boldsymbol{h}}$. Therefore, the heatmaps data, $\widetilde{\boldsymbol{h}}$ and $\widehat{\boldsymbol{h}}$, are correlated with each other. This motivates us to adopt a two-channel ConvLSTM model, with first channel as heatmap data $\hat{h}_\omega \in \widehat{\boldsymbol{h}}$ and second channel as FLM heatmap data, $\tilde{h}_\omega \in \widetilde{\boldsymbol{h}}$. This two-channel build-up maintains the correlations between different channel data, which is able to accurately and simultaneously predict the heatmaps for analyzing the ridesharing swarms and FLM zones. Specifically, the input data is a sequence of $\mathrm{S} \times \mathrm{S}$, two-channel images, or a series of tensors, $X_t \epsilon R^{S \times S \times 2}$. We validate the performance of the two-channel ConvLSTM learning model by the case study below.

2.4 CASE STUDY AND RESULTS

This case study validates the capability of the proposed analysis approach based on the real field data. More exactly, we will examine the discretization approach as well as its capability in finding ridesharing swarms, and inferring and prediction of passenger demands in different levels of variations.

The case study is built upon the testbed consisting of the transit and ridesharing services data in the city of Chengdu. Chengdu is a major city in China and has a population of 7.8 million. As shown in Figure 27 , the study area is in the second ring region of the city and covers a square region with 5 miles edge length. The ridesharing service data is provided by DiDiChuxing Gaia open dataset (https://gaia.didichuxing.com). It involves about 0.2 million trips made by DiDi ridesharing services per day from November 1 to November 30, 2016. The profile of the data includes ridesharing vehicle trajectory GPS data, which is updated 2 – 4 seconds, and ridesharing order request information, which records pick-up and drop-off timestamps and locations. The public transit data is collected by the website of Moovit app (https://moovitapp.com/). It covers all bus lines and subway lines information in Chengdu city. For each line, the profile of the data includes station names, station locations, operation times and transit vehicles' arrival time at each station. Please note that the real time trajectory data of transit vehicles is unavailable. This study will use the schedule and station location data to format each trip. The operating time of transit services varies from line to line, but most of them start at 6 AM and end at 8 PM to 10 PM. There are a total 1226 transit stations distributed within the study area and, in total, 246 transit lines passing through the study area. As the study area is 5×5 mi^2, there are about 7 transit stations per mile on average. The case study is run on a DELL Precision 3630 Tower with 3.60GHz of Intel Core i9-9900k CPU and 16 GB RAM in a Windows environment.

We briefly introduce the procedure the case study follows. The case study will first determine the optimal discretization (τ^*, κ^*) for DiDi trip data and transit line data in 3D space. The heatmap $\boldsymbol{h} = \{h_i, i \in I\}$ is then generated based on the optimal discretization for each interval $i \in I$. Then, we aggregate the heatmap to find RS zones and reveal the locations for new transit lines and stations (with microtransit services) (see Section 2.3.2 Searching Ridesharing Swarm). Next, we explore the FLM zones by analyzing the heatmaps $\widehat{\boldsymbol{h}} = \{\hat{h}_\omega, \omega \in \mathcal{W}\}$ and ridesharing OD information (see Section 2.3.3 Searching First and Last Mile (FLM) gap). Finally, the results

are fed into a ConvLSTM network to predict the dynamics of the FLM demand pattern (see Section 2.3.4. Learning Spatiotemporal Service Gaps). Last, we also analyze land use patterns of RS zones and FLM zones to validate our findings.

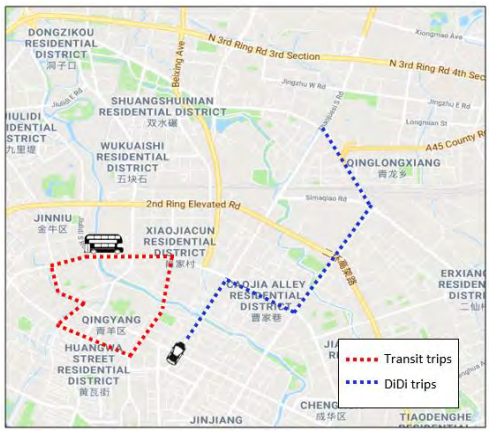


Figure 27 Chengdu second ring

2.4.1 Establishing 3D Discretization

We first establish the 3D discretization (τ^*, κ^*) for the trip data involved in this case study. To do that, we need to determine the optimal time interval τ^*. The solution time of the program $\boldsymbol{P_1}$ is around 20 minutes. And the optimal solution we obtained is $\tau^* = 90s$, with which we averaged about 30 coordinates of each vehicle to locate it in a time interval. Next, we explore the optimal κ^* through the program $\boldsymbol{P_2}$, which needs to pre-determine the parameters $(\underline{\eta}, \overline{\eta}, \underline{\mu}, \overline{\mu})$. By setting $\underline{\eta} = 0.1, \overline{\eta} = 0.9, \underline{\mu} = 0.4, \overline{\mu} = 0.6$, it took program $\boldsymbol{P_2}$ around 2500 seconds to find the optimal number of the pixels, $\kappa^* = 28^2$, which indicates a pixel size: 325m×325m for each cube.

According to our discussion in Section 3.1.2, we justify the selection of the parameters $\underline{\eta}$, and $\underline{\mu}$ for this case study by doing the sensitivity analysis on the length of the UI, i.e., the length of $\left(\underline{\mu} - \underline{\eta}\right)$. Mainly, with a given time interval τ^*, we test the performance of the 3D discretization under each of the four UIs shown in Table 12, where the UI varies from 0.1 to 0.4 and each corresponds to a set of parameters selection.

Table 12 Sensitivity analysis of UIs

UI	$(\underline{\eta}, \overline{\eta}, \underline{\mu}, \overline{\mu})$	κ^*	Total noise ratio
0.40	(0.05, 0.95, 0.45, 0.55)	34^2	0.62
0.30	**(0.10, 0.90, 0.40, 0.60)**	$\mathbf{28^2}$	**0.45**
0.20	(0.15, 0.85, 0.35, 0.65)	16^2	0.37
0.10	(0.20, 0.80, 0.30, 0.60)	10^2	0.13

Program $\boldsymbol{P_2}$ is run under each UI and generates the optimal solution κ^* shown in Table 12. Upon each optimal discretization scheme (τ^*, κ^*), the heatmaps $\boldsymbol{h} = \{h_i, i \in I\}$ were generated

and shown as the examples in Figure 28, in which the region completely dominated by transit or ridesharing is colored red and white respectively, but the regions where neither transit nor ridesharing services show are in black. Accordingly, the complementary areas are in either red (transit service-dominant) or white color (ridesharing service-dominant) and competitive regions and unidentifiable pixels (see Definition 3) are orange with different intensity. In addition, Figure 28(a) illustrates the solution of a 3D discretization, while Figure 28(b) and (c), respectively, present a heatmap (10:00:00 - 10:01:30) for the solution with $\kappa^* = 28^2$ and $\kappa^* = 16^2$ with the UI equal to 0.3 and 0.2. Note that there are too many cubes to be clearly outlined in Figure 6(a). We instead show the height of cubes, τ^*, along time dimension.

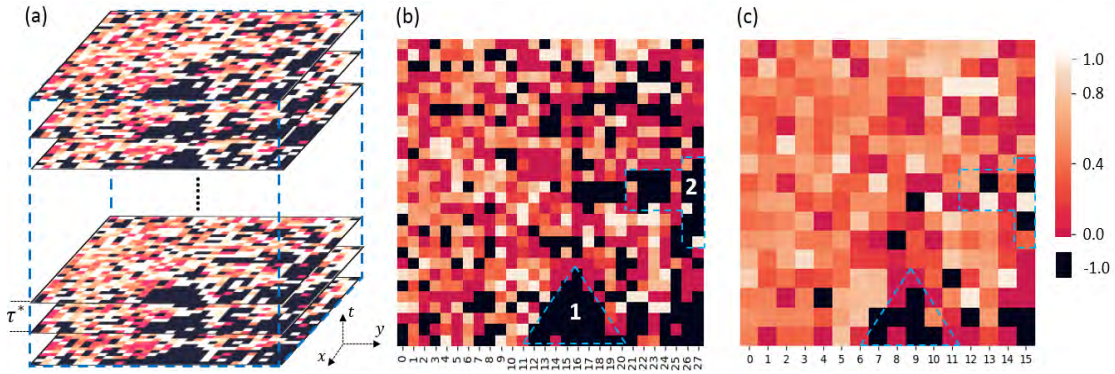


Figure 28 (a) Discretization of transit and ridesharing trip data in 3D space. Land-use analysis on heatmaps with different pixel numbers to evaluate resolution. (b) 10:00:00-10:01:30 heatmap with optimal pixel number 28^2 (0.3 UI). (c) 10:00:00-10:01:30 heatmap with optimal pixel number 16^2 (0.2 UI).

We evaluated the merits of the heatmaps in terms of the land-use pattern and the ratio of noise pixels over all pixels in a 3D discretization solution. We are more interested in a 3D discretization that has a smaller ratio of noise pixels, because the noise pixels don't offer much insightful information about the service gaps for both service modes. The results in Table 12 indicate that the UI of 0.3 is better than the UI of 0.4 since the ratio of noise pixels is decreased from 0.62 to 0.45. This ratio can be further reduced from 0.45 to 0.13 as the UI decreases from 0.3 to 0.1 but with the significant loss of the resolution. Specifically, we conducted the land-use analysis to examine the resolution of the heatmaps shown in Figure 28(b) and (c). The region 1 and region 2 (marked out by the dashed outline) are chosen as the benchmark regions. The actual land-use analysis shows that region 1 is industrial land with a large waste disposal plant, several major intercity railway lines and railway companies. Region 2 mainly consists of parks, intercity highways and highway interchange junctions. These results indicate that both regions have few mobility needs and limited attractiveness for transit and ridesharing services. We next take a look at the heatmap results. The contours of these two regions are clearly outlined in Figure 28(b) and they are mainly in black (i.e., no transport services provided). This is consistent with the actual land-use analysis. But Figure 28(c) does not demonstrate the same quality of the resolution. Thus, we conclude that reducing UI from 0.2 to 0.1 significantly compromises the

resolution of heatmaps. The above analysis confirms our best choice of UI (= 0.3) for this case study.

2.4.2 Finding Ridesharing Swarm Zones

According to the approach developed in Section 2.3.2, we project the heatmaps $h_i, i \in I$ to a 2D spatial plane and then get the aggregated heatmap H shown in Figure 29. Built upon H, we search the ridesharing swarms (RS) on the heatmap H. The results are shown in Figure 29 (a), in which six major RS zones are marked by the dashed line cycles. There are also some small RS zones scattered over the heatmap H. Combined with the land-use data, these major RS zones involve a commercial region (marked by a yellow circle on the left side), a residential region (marked by a blue circle), a mixed region (marked by a purple circle) including residence community, office buildings and shopping centers, and a big theme park (green circle) in Chengdu City.

The land-use results indicate that these RS zones do attract/generate significant demand. On the other hand, the heatmap results show that ridesharing services dominate these mobility services. These observations together reveal potential deficiency of transit services in either coverage or flexibility in those RS zones. In order to provide valuable suggestions for improving the transit services in these RS zones, we further investigate the number of ridesharing trips between these major RS zones and show the results in Table 13. It is observed that a great number of daily ridesharing trips occur between the RS zones 1 and 3 (2531.4 per day), RS zones 3 and 4 (2762.8 per day), RS zones 3 and 5 (2878.5 per day), and RS zones 4 and 5 (7617.8 per day). These results are consistent with the land use features of these RS zones, which are either large commercial areas or residential communities near the major metro hub, thus attracting significant and stable demand. For those RS zones, we would suggest adding new transit stations in these RS zones and additional routes between them (see the blue dashed lines, Figure 29 (b)) so that public transit can catch some of the demand currently using

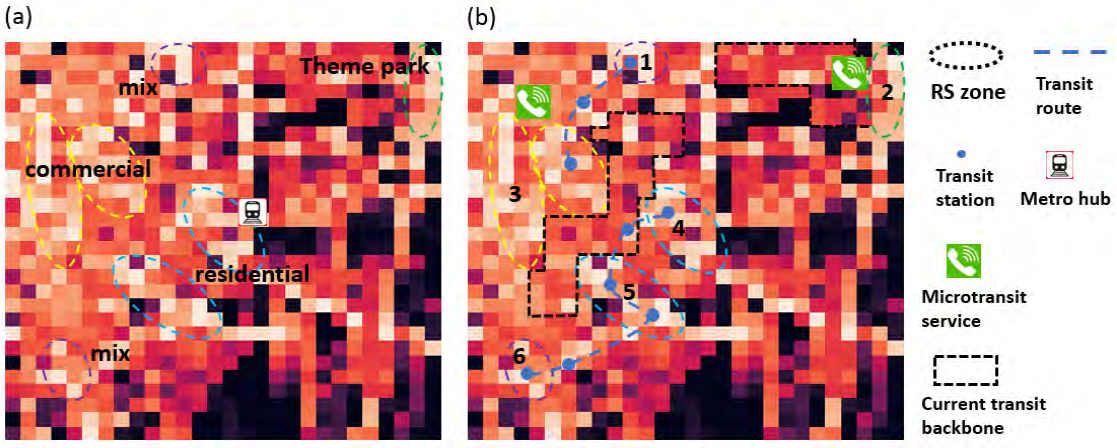


Figure 29 (a) Identified RS zones on H with land use. (b) Potential new transit lines, stations and microtransit services.

ridesharing services. The mean number of daily trips and standard deviation also provide hints on the design of the transit schedule, i.e., line frequency.

On the other hand, this study would suggest microtransit service for the zones that are characterized by high demand in a large area with low inside transit coverage. For example, our data analysis observed that the number of daily demands to zones 3 and 2 are around 36,000 and 15,000 trips per day, which respectively counts for 19.5% and 8% of total ridesharing trips of the study area. Moreover, these zones are large, e.g., about 2.69 mi^2 for zone 3, with low inside transit coverage, but surrounded by enough transit services. These features can be observed from the heatmap. Figure 29 (b) uses a black dashed line to outline the areas covered by a transit backbone (i.e., they are the red transit dominant pixels and orange competitive pixels). These features indicate that integrating on-demand service, e.g., microtransit, ridesharing, or mini-bus services into the current transit system will improve the mobility of these RS zones. Please note that the approaches of this study help identify the potential areas that need better transit or hybrid mobility services. To further judge which guidance fits best to these RS zones and how to implement these suggestions properly is out of the scope of this study.

Table 13 Number of trips per day between RS zones

RS zones pair	Mean/day	Standard Deviation	RS zones pair	Mean/day	Standard Deviation
(1,2)	250.9	17.6	(2,6)	5.0	1.7
(1,3)	**2531.4**	**243.5**	**(3,4)**	**2762.8**	**159.5**
(1,4)	463.7	76.2	**(3,5)**	**2878.5**	**396.2**
(1,5)	215.6	49.7	(3,6)	431.7	76.2
(1,6)	26.8	7.5	**(4,5)**	**7617.8**	**709.2**
(2,3)	172.0	15.1	(4,6)	131.0	12.7
(2,4)	655.9	94.8	**(5,6)**	**850.6**	**103.9**
(2,5)	333.6	40.8			

2.4.2 Inferring FLM Zones

To explore the potential FLM zones, we averaged every 40 heatmap $h_i, i \in I$, each of which

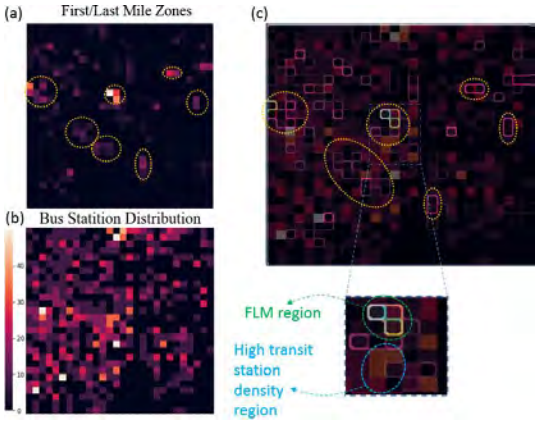


Figure 30 Identified FLM zones (a) Identified FLM zones (b) transit station distribution map. (c) Overlap of FLM zones and transit services distribution

represents trip information in 90 seconds , to generate the aggregated heatmap $\hat{h}_\omega, \omega \in \mathcal{W}$, each of which contains 1-hour trip information. Upon $\hat{\boldsymbol{h}} = \{\hat{h}_\omega, \omega \in \mathcal{W}\}$, we search those sandwich areas and then identify the FLM zones by integrating the OD information. Figure 30(a) presents an example of the results, in which the FLM zones are marked by the yellow dashed cycle. The whiter the pixel, the greater number of prone-FLM orders the pixel has. Due to the lack of intermodal trip data, we validate these FLM zones by integrating bus station data and land-use data. Specifically, we agree with the criterion that an FLM zone often occurs in areas with lower transit density than the surrounding areas. This situation likely pushes passenger demand to use ridesharing or other traffic modes since the demand needs a long walking distance to approach nearby transit backbone lines. With this observation, we compared the transit service density within the FLM zones with surrounding areas to validate the FLM zones. Figure 30(b) presents the map of bus station distribution. Each white pixel includes around 50 bus stations, while the black pixel does not possess bus stations. By integrating Figure 30(a) and Figure 30(b), we obtain Figure 30(c), in which the color of the pixel contour represents the heat of FLM intensity and the pixel color represents the density of bus stations therein. We observed that all the FLM pixels are quite dark inside but bright in the surrounding contour. Therefore, the FLM zones we found indeed have lower transit service density than the surrounding areas, which is consistent with our criteria. This observation validates the FLM zones we found.

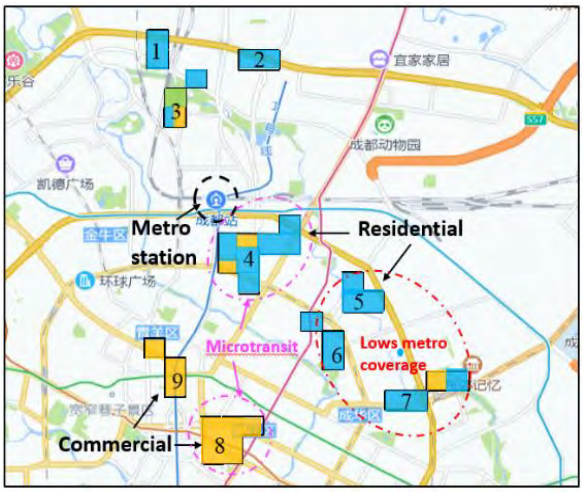


Figure 31 Land use validation for top FLM zones. Commercial area ( ), residential area ( ).

Next, we consider that an attractive FLM area should generate enough demand. This motivates us to validate the FLM zones by examining the land-use of the study region. We selected the top nine potential FLM zones found in this case study and checked the corresponding land use in Gaode Map. The results are shown in Figure 31, in which commercial areas (building offices/malls/hotels) are marked by yellow, while residential areas are blue. We can see that all identified FLM zones are either commercial or residential areas, which have a high population density. Specifically, zones (1, 3, 4, 8, 9) are those areas located around a metro line or within suburban areas. In addition, the map data indicate that these districts are quite large but with sparse transit service. There are, indeed, the FLM zones that are not well connected to the existing backbone transit lines. Here, shared mobility service such as ridesharing or microtransit services is a good complement to the backbone transit lines. We also noticed that Zones (2, 5, 6, 7) with low transit service are far away from a metro line. They are typical transit deficiency zones, where new transit routes are needed to make up the service gaps. Overall, the case study shows our approach works efficiently to find FLM zones and will provide valuable information to refine current transit networks.

2.4.3 Predicting FLM

We next demonstrate the performance of the machine learning model to predict the spatiotemporal service gaps. Specifically, we implement the ConvLSTM using the framework proposed by (Xingjian et al., 2015). Specifically, the model is implemented using the Keras API. It consists of five layers. The first and third layers are ConvLSTM with 120 size 2×2 kernels and 80 size 3×3 kernels, respectively. The second and fourth layers are batch normalization. Finally, we used $2 \times 2 \times 2$ kernels to get output shape. Table 14 provides details of the model parameters.

Table 14 Description of the ConvLSTM architecture used in the study

Layer (type)	**Output Shape**	**Param Number**
ConvLSTM2D	(None, None, 28, 28, 120)	234720
Batch Normalization	(None, None, 28, 28, 120)	480
ConvLSTM2D	(None, None, 28, 28, 80)	576320
Batch Normalization	(None, None, 28, 28, 80)	320
Conv3D	(None, None, 28, 28, 2)	1282
Total params	813122	
Trainable params	812722	
Non-trainable params	400	

We use Mean Average Percentage Error (MAPE) and Location Prediction Error (LPE) to evaluate the performance of ConvLSTM network, which are defined as follows:

$$MAPE = \frac{1}{\kappa}\sum_{k=1}^{\kappa}\frac{|\hat{y}_{i+1}^{k} - y_{i+1}^{k}|}{y_{i+1}^{k}} \tag{14}$$

$$LPE = \frac{1}{\kappa}\sum_{k=1}^{\kappa}|\hat{\delta}_{i+1}^{k} - \delta_{i+1}^{k}| \tag{15}$$

where $\hat{y}_{i+1}^{k}$, y_{i+1}^{k} are the prediction and real value of pixel k for time interval $i+1$. $\hat{\delta}_{i+1}^{k}$, δ_{i+1}^{k} take 0-1 values, indicates whether pixel k is FLM zone. κ is the total number of pixels in an image. The goal of FLM prediction is to tell where there are potential FLM zones in the near future to help planning and scheduling of FLM micro-bus service. Therefore, LPE is adopted to evaluate the performance of FLM prediction. Note that we focus on the location prediction accuracy of FLM zones.

The training data consists of $\hat{h}_{\omega} \in \widehat{\boldsymbol{h}}$ as first channel input and corresponding $\tilde{h}_{\omega} \in \widetilde{\boldsymbol{h}}$ as second channel input. The output is the two-channel prediction: $\hat{h}_{\omega}$ and $\tilde{h}_{\omega}$ in the next time interval. The model is trained with the first three weeks of data. The last week of data is used for validation. The model achieves 28.42% (MAPE) for the first channel and 4.22% (LPE) for the second channel. Therefore, the model exhibits high accuracy in predicting the locations of FLM zones, which provides promising radar for the FLM demand and the corresponding ridesharing or microtransit services.

Moreover, the prediction results also indicate the time-variant characteristics of the FLM demand. For example, Figure 32 (a) to (c) indicates the interesting dynamics of FLM demand during the period covering morning peak hour (8:00-9:00), secondary-peak hour (9:00-10:00) and evening peak hour (18:00-19:00). More exactly, we noticed that the commercial zone (big

shopping centers) on the upper-left corner has lower FLM demand during 8:00 to 9:00. This is reasonable since most of shopping malls are closed at that time. Moreover, the FLM demand in the residential zone, i.e. r_1, during morning and evening peak hours is higher than that during the secondary-peak hour. This observation indicates fewer work trips occurring during secondary-peak hour. We also notice that the residential zone r_2 consistently has high FLM demand. This is because the new metro hub nearby attracts a stable and significant FLM demand, which calls for the transit service coverage in the area of r_2 by either creating new transit station (inflexible service) or providing microtransit (flexible service). These interesting findings uncover the dynamics of potential transit demand and benefit future hybrid transit system development.

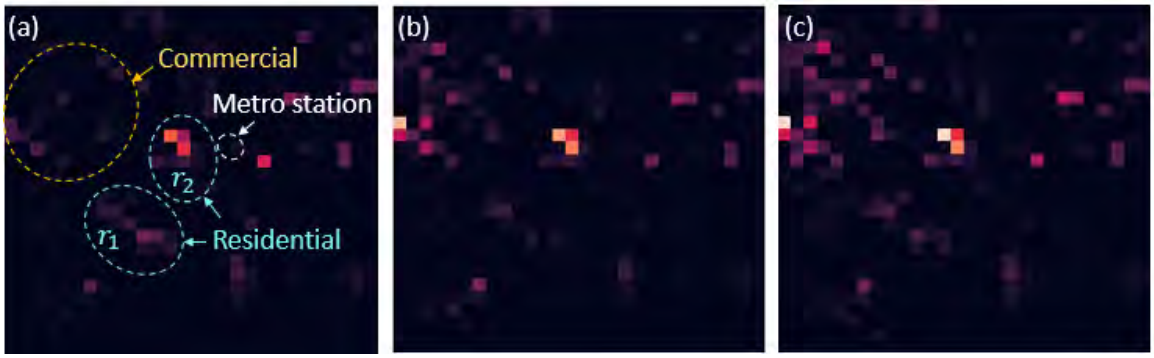


Figure 32 Prediction of time-vary FLM heatmap $\mathrm{h}_{\mathrm{FLM}}$. FLM heatmap of (a) 8:00-9:00, (b) 9:00-10:00, (c) 18:00-19:00. Whiter pixel indicates a higher possibility of FLM *demand*.

2.5 CONCLUSION

This study developed an innovative approach to analyze transit and ridesharing trip data for discovering mobility needs with different levels of dynamics to support the development of hybrid urban public transport systems involving both transit and on-demand services. More exactly, we think the complicated competitive and complementary relationship between transit and ridesharing trips can reflect the deficiency in the flexibility and coverage of the existing transit services and provide valuable guidelines to integrate transit service with on-demand mobility services. To uncover this hidden knowledge in the trip data, we developed the novel 3D trip data analysis approach, which first meshes trip data into an optimal 3D discretization with uniform cube size, and then collects information from each cube to form the heatmaps. Built upon the heatmaps, we examined the ridesharing swarm zones, which help discover the potential deficiency of the flexibility and coverage of existing transit services, and then provide suggestions for either refining the existing transit network and schedules or integrating ridesharing or microtransit services. Next, we developed a new approach to infer the first/last mile zones by discovering the "sandwich" patterns on the aggregated heatmaps. The identified first/last mile zones provide guidelines for integrating microtransit and/or ridesharing services. Last, feeding the heatmaps into a two-channel ConvLSTM model, this analysis predicts the

dynamics of the spatiotemporal service gaps, which will help the hybrid urban public transport system make strategic plans for involving both public transit and on-demand services. A case study conducted for the second ring region of Chengdu, China validates the effectiveness and capability of our analysis approach. This study is our first attempt to discover transit service gaps by integrating transit and ridesharing trip data. The analysis can be extended to involving other mobility modes that transit competes with such as bike-sharing and private auto. Even though this paper primarily used trip data, the developed approach can incorporate other data, such as demographic and weather data to reach a more comprehensive understanding of transit service gaps and improve the prediction accuracy.

2.6 REFERENCE LIST

1. Aldaihani, M.M., Quadrifoglio, L., Dessouky, M.M., Hall, R., 2004. Network design for a grid hybrid transit service. Transp. Res. Part A Policy Pract. 38, 511–530.
2. Alemi, F., Rodier, C.J., University of California, D.I. of T.S., 2016. Simulation of ridesourcing using agent-based demand and supply regional models : potential market demand for first-mile transit travel and reduction in vehicle miles traveled in the San Francisco Bay Area.
3. Berrada, J., Andreasson, I., Burghout, W., Leurent, F., n.d. Demand modelling of autonomous shared taxis mixed with scheduled transit. Transp. Res. Board 98th Annu. Meet.
4. Boarnet, M.G., Giuliano, G., Hou, Y., Shin, E.J., 2017. First/last mile transit access as an equity planning issue. Transp. Res. Part A Policy Pract. 103, 296–310.
5. Boyle, D.K., 2006. Fixed-route transit ridership forecasting and service planning methods. Transportation Research Board.
6. Chatterjee, A., Venigalla, M.M., 2004. Travel demand forecasting for urban transportation planning. Handb. Transp. Eng. 1.
7. Curtis, T., Merritt, M., Chen, C., Perlmutter, D., Berez, D., Ellis, B., 2019. Partnerships Between Transit Agencies and Transportation Network Companies. Transportation Research Board, Washington, D.C. https://doi.org/10.17226/25425
8. Davidson, A., Peters, J., Brakewood, C., n.d. Interactive Travel Modes: Uber, Transit, and Mobility in New York City. Transp. Res. Board 96th Annu. Meet.
9. Dechter, R., Pearl, J., 1985. Generalized best-first search strategies and the optimality of A. J. ACM 32, 505–536.
10. Faghih, S.S., Safikhani, A., Moghimi, B., Kamga, C., 2019. Predicting Short-Term Uber Demand in New York City Using Spatiotemporal Modeling. J. Comput. Civ. Eng. 33, 5019002.
11. Fang, Z., Cheng, Q., Jia, R., Liu, Z., 2018. Urban rail transit demand analysis and prediction: A review of recent studies, in: International Conference on Intelligent Interactive Multimedia Systems and Services. pp. 300–309.
12. Farhan, J., Chen, T.D., Zhang, Z., 2018. Leveraging Shared Autonomous Electric Vehicles for First/Last Mile Mobility. Transp. Res. Board 97th Annu. Meet.
13. Feigon, S., Murphy, C., 2016. Shared mobility and the transformation of public transit.
14. Fu, L., 2002. Planning and design of flex-route transit services. Transp. Res. Rec. 1791, 59–66.
15. Hall, J.D., Palsson, C., Price, J., 2018. Is Uber a substitute or complement for public transit? J.

Urban Econ. 108, 36–50.
16. Hashemian, H., 2002. Using GIS to Assess Demographic and Land Use Characteristics on Local Transit Services, in: Seventh TRB Conference on the Application of Transportation Planning MethodsTransportation Research Board; Commonwealth of Massachusetts, Executive Office of Transportation and Construction; and Boston Metropolitan Planning Organization.
17. Hill, S., 2018. Ridesharing Versus Public Transit [WWW Document]. Am. Prospect. URL url: http://prospect.org/article/ridesharing-versus-public-transit
18. Huang, H., 1996. The land-use impacts of urban rail transit systems. J. Plan. Lit. 11, 17–30.
19. Jaller, M., Pourrahmani, E., Rodier, C., Bischoff, J., 2019. Simulation-Optimization Framework to Evaluate a Sustainable First Mile Transit Access Program Using Shared Mobility. Transp. Res. Board 98th Annu. Meet.
20. Jun, M.-J., Choi, K., Jeong, J.-E., Kwon, K.-H., Kim, H.-J., 2015. Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul. J. Transp. Geogr. 48, 30–40.
21. Koffman, D., 2004. Operational experiences with flexible transit services. Transp. Res. Board Natl. Acad. TCRP Synth.
22. Liu, L., Qiu, Z., Li, G., Wang, Q., Ouyang, W., Lin, L., 2019. Contextualized spatial--temporal network for taxi origin-destination demand prediction. IEEE Trans. Intell. Transp. Syst. 20, 3875–3887.
23. Nazem, M., Trépanier, M., Morency, C., 2011. Demographic analysis of route choice for public transit. Transp. Res. Rec. 2217, 71–78.
24. Nourbakhsh, S.M., Ouyang, Y., 2012. A structured flexible transit system for low demand areas. Transp. Res. Part B Methodol. 46, 204–216.
25. Pinto, H.K.R. de F., Hyland, M.F., Verbas, İ.Ö., Mahmassani, H.S., 2018. Integrated Mode Choice and Dynamic Traveler Assignment-Simulation Framework to Assess the Impact of a Suburban First-Mile Shared Autonomous Vehicle Fleet Service on Transit Demand. Transp. Res. Board 97th Annu. Meet.
26. Potts, J.F., Marshall, M.A., Crockett, E.C., Washington, J., 2010. A guide for planning and operating flexible public transportation services.
27. Qiu, F., Li, W., Zhang, J., 2014. A dynamic station strategy to improve the performance of flex-route transit services. Transp. Res. Part C Emerg. Technol. 48, 229–240.
28. Quadrifoglio, L., Li, X., 2009. A methodology to derive the critical demand density for designing and operating feeder transit services. Transp. Res. Part B Methodol. 43, 922–935.
29. Roberts, R.A., 1985. Analysis of demographic trends and travel patterns: Implications for the future of the Portland transit market. Transp. Res. Rec. J. Transp. Board 1–8.
30. Shen, Y., Zhang, H., Zhao, J., 2017. Embedding Autonomous Vehicle Sharing in Public Transit System: An Example of Last-Mile Problem. Transp. Res. Board 96th Annu. Meet.
31. Stiglic, M., Agatz, N., Savelsbergh, M., Gradisar, M., 2018. Enhancing urban mobility: Integrating ride-sharing and public transit. Comput. Oper. Res. 90, 12–21.
32. Sung, H., Choi, K., Lee, S., Cheon, S., 2014. Exploring the impacts of land use by service coverage and station-level accessibility on rail transit ridership. J. Transp. Geogr. 36, 134–140.
33. Vakayil, A., Gruel, W., Samaranayake, S., n.d. Integrating Shared-Vehicle Mobility-on-

Demand Systems with Public Transit. Transp. Res. Board 96th Annu. Meet.
34. Velaga, N.R., Nelson, J.D., Wright, S.D., Farrington, J.H., 2012. The potential role of flexible transport services in enhancing rural public transport provision. J. Public Transp. 15, 7.
35. Xingjian, S.H.I., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting, in: Advances in Neural Information Processing Systems. pp. 802–810.
36. Xu, H., Pang, J.-S., Ordóñez, F., Dessouky, M., 2015. Complementarity models for traffic equilibrium with ridesharing. Transp. Res. Part B Methodol. 81, 161–182.
37. Xu, J., Rahmatizadeh, R., Bölöni, L., Turgut, D., 2017. Real-time prediction of taxi demand using recurrent neural networks. IEEE Trans. Intell. Transp. Syst. 19, 2572–2581.
38. Yan, X., Levine, J., Zhao, X., 2018. Integrating ridesourcing services with public transit: An evaluation of traveler responses combining revealed and stated preference data. Transp. Res. Part C Emerg. Technol. https://doi.org/10.1016/J.TRC.2018.07.029
39. Zhang, J., Zheng, Y., Qi, D., 2016. Deep spatio-temporal residual networks for citywide crowd flows prediction. arXiv Prepr. arXiv1610.00081.
40. Zhang, K., Liu, Z., Zheng, L., 2019. Short-term prediction of passenger demand in multi-zone level: Temporal convolutional neural network with multi-task learning. IEEE Trans. Intell. Transp. Syst. 21, 1480–1490.
41. Zhou, X., Shen, Y., Zhu, Y., Huang, L., 2018. Predicting multi-step citywide passenger demands using attention-based neural networks, in: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. pp. 736–744.

3.0 RECOMMENDATIONS

This study can be extended in a number of directions. First, the findings derived from Task 1 and 2 may be limited to the dataset collected from different cities, the second ring area of Chengdu and the Orlando metropolitan area. Future research can be extended to a comprehensive case study based on a complete dataset in a city and investigating the transferability of the findings. Task 1 and Task 2 can be integrated in this future work. More exactly, the approaches developed by Task 2 will first identify the areas with the service gaps in the temporal-spatial transit service network. Built upon the potential areas found in Task 2, the approaches in Task 1 can further analyze the demand characteristics in those potential areas so that we are able to provide proper suggestions for integrating microtransit and/or ridesharing into the existing transit network or for adding/adjusting the transit network and service.

Second, the analysis for the supply market is our first attempt to discover transit service gaps by integrating transit and ridesharing trip data. This analysis can be extended to incorporate other mobility data and modes that transit competes for demand such as bike-sharing and private auto. Finally, the supply market analysis primarily used trip data. The developed approach can also incorporate other data, such as demographic and meteorological data to reach a more comprehensive understanding of transit service gaps and improve the prediction accuracy.

4.0 APPENDICES

4.1 Appendix A – Acronyms, abbreviations, etc.

Acronyms	Definition
ADA	Americans with Disabilities Act
AIC	Akaike's Information Criterion
AMoD	Autonomous Mobility-On-Demand
aTaxis	Autonomous Taxis
AV	Autonomous Vehicle
BART	Bay Area Rapid Transit
BIC	Bayesian Information Criterion
BRT	Bus Rapid Transit
CBG	Census Block Group
CBSA	Core-Based Statistical Area
CGT	Connetics Transportation Group
ConvLSTM	Convolutional Long Short Term Memory
CTA	Chicago Transit Authority
EPA	Environment Protection Agency
FLM	First and last mile
GIS	Geographic Information Systems
GTFS	General Transit Feed Specification
HH	Household
HU	Housing Units
KNR	Kiss-and-Ride
LPE	Location Prediction Error
LRT	Log-Likelihood Ratio Test
LSTM	Long Short Term Memory
MAPE	Mean Average Percentage Error
MNL	Multinomial Logit
MSA	Metropolitan Statistical Areas
PNR	Park-and-Ride
RP	Revealed Preference
RSG	Resource Systems Group
SAEV	Shared Autonomous Electric Vehicles

Acronyms	Definition
SAV	Shared Autonomous Vehicles
SC	Schwarz Criterion
SLD	Smart Location Database
SP	Stated Preference
TNC	Transportation Network Company
VMT	Vehicle Miles Traveled

4.2 Appendix B – Associated websites, data, etc., produced

Table 15 Model Results for Access Mode

Base Category	Parameter	Wheelchair	Micro-mobility	Carpool	TNC or Taxi	Drove Alone
	Intercept	-6.5879 (0.0023)	-3.6756 (0.0012)	-3.7920 (0.0019)	-5.2544 (0.0033)	-2.9104 (0.0011)
Access Length	**Distance from the origin to transit stop (mile)**		1.3802 (0.0004)	1.0966 (0.0004)	1.2077 (0.0009)	0.8382 (0.0011)
Destination Place	**Airport**	-0.6220 (0.0765)	0.6333 (0.0141)	1.6289 (0.0077)	4.5079 (0.0080)	0.7242 (0.0208)
	Medical, Hospital	-1.7809 (0.0145)	0.1199 (0.0061)	0.4373 (0.0057)	-0.7863 (0.0345)	0.1350 (0.0102)
	Sporting Events	-0.9363 (0.0944)	0.9780 (0.0144)	3.4428 (0.0080)	-3.0877 (0.4182)	2.3677 (0.0155)
	University/college	-0.3235 (0.0214)	0.7579 (0.0050)	0.1250 (0.0060)	1.0043 (0.0122)	1.0782 (0.0064)
Origin Place	**Shopping**	-0.3878 (0.0088)	-0.5704 (0.0085)	-0.6115 (0.0071)	0.3203 (0.0171)	-0.3722 (0.0111)
	Social Visit	0.7676 (0.0057)	0.2216 (0.0048)	0.4953 (0.0041)	0.2782 (0.0125)	-0.2334 (0.0104)
Transfer	**Number of transfers from the origin**	-0.1726 (0.0033)	-0.2860 (0.0021)	-0.2565 (0.0019)	-0.2274 (0.0060)	-0.4630 (0.0036)
	Number of transfers to the destination	-0.1730 (0.0035)	-0.3844 (0.0022)	-0.2476 (0.0019)	-0.1449 (0.0052)	-0.4698 (0.0035)
Two-way trip	**Trip in the Opposite Direction-Yes**	0.0474 (0.0035)	0.1887 (0.0016)	0.0287 (0.0016)	-0.3186 (0.0053)	0.2456 (0.0024)
Visitor	**Visitor-Yes**	-0.6948 (0.0138)	-0.8573 (0.0061)	0.0533 (0.0039)	-0.8857 (0.0102)	-0.4065 (0.0081)
Time Period	**Midday**	0.0369 (0.0033)	-0.1284 (0.0021)	-0.2435 (0.0020)	-0.3179 (0.0061	-0.2294 (0.0033)
	Evening	-0.2170 (0.0071)	-0.0934 (0.0035)	-0.2140 (0.0033)	-0.7531 (0.0114)	-0.6177 (0.0069)
Month (refernce -January)	**December**	0.4852 (0.0070)	0.2096 (0.0039)	0.0561 (0.0038)	-0.1235 (0.0169)	-0.2026 (0.0081)
	February	0.4805 (0.0035)	-0.1917 (0.0024)	-0.1677 (0.0023)	-0.0660 (0.0081)	0.1235 (0.0036)

Base Category	Parameter	Wheelchair	Micro-mobility	Carpool	TNC or Taxi	Drove Alone
	November	-0.4131 (0.0284)	0.1986 (0.0137)	0.1971 (0.0132)	0.9348 (0.0364)	-0.1984 (0.0307)
Age	**Young Adults (18-34 years old)**	-0.2057 (0.0054)	-0.2083 (0.0019)	-0.1087 (0.0017)	-0.2024 (0.0052)	-0.2400 (0.0029)
Disability	**Disability-Yes**	5.6846 (0.0024)	-0.1091 (0.0035)	-0.1853 (0.0040)	-0.4964 (0.0146)	-0.1382 (0.0068)
Driver License	**Driver License-Yes**	-0.3566 (0.0043)	-0.1025 (0.0017)	0.1838 (0.0015)	0.2740 (0.0039)	0.7187 (0.0022)
Ethnicity	**African American**	0.0350 (0.0034)	-0.4618 (0.0020)	-0.1985 (0.0018)	-0.3017 (0.0058)	-0.4848 (0.0033)
	Asian	1.0739 (0.0166)	-0.8116 (0.0132)	-0.1800 (0.0085)	-0.3127 (0.0220)	-0.1952 (0.0135)
	Hispanic	-0.4542 (0.0063)	-0.4058 (0.0025)	-0.1067 (0.0023)	0.1160 (0.0059)	-0.2807 (0.0038)
Gender	**Male**	-0.1961 (0.0032	1.1017 (0.0015)	-0.0648 (0.0016)	-0.3152 (0.0051)	-0.1626 (0.0027)
Number of Vehicles	**Number of Vehicles (4-7)**	-0.3735 (0.0293)	0.2450 (0.0082)	0.7470 (0.0062)		0.6076 (0.0104)
HH Income	**Middle Income ($20K-$50K)**	-0.4078 (0.0048)	0.2643 (0.0018)	0.1236 (0.0018)	-0.1562 (0.0051)	0.1743 (0.0030)
	High Income ($50K-$100K)	-0.5003 (0.0142)	0.4240 (0.0033)	0.7081 (0.0029)	0.4793 (0.0075)	1.3374 (0.0035)
	Very High Income (More than $100K)	-1.0835 (0.0639)	1.8597 (0.0067)	4.7172 (0.0037)	0.9611 (0.0317)	2.6380 (0.0092)
-	**Employment and household entropy**	-0.2846 (0.0036)	0.0218 (0.0020)	-0.0740 (0.0018)	-0.3697 (0.0054)	-0.1725 (0.0030)
-	**Gross education(8-tier) employment density (jobs/acre) on unprotected land**	-0.0089 (0.0002)	0.0059 (0.0001)	0.0032 (0.0002)	0.0051 (0.0003)	0.0046 (0.0003)
-	**Gross entertainment (5-tier) employment density (jobs/acre) on unprotected land**	0.0875 (0.0009)	-0.0246 (0.0007)	-0.0387 (0.0006)	0.0220 (0.0015)	-0.0578 (0.0012)
-	**Gross industrial (5-tier) employment density (jobs/acre) on unprotected land**	0.0588 (0.0012)	0.0728 (0.0009)	-0.0349 (0.0008)	0.0837 (0.0025)	-0.0138 (0.0015)
-	**Gross office (8-tier) employment density (jobs/acre) on unprotected land**	-0.0479 (0.0007)	0.0431 (0.0006)	0.0104 (0.0004)		-0.0074 (0.0009)
-	**Gross residential density (HU/acre) on unprotected land**	-0.0672 (0.0007)	-0.0141 (0.0004)	-0.0396 (0.0004)	-0.0144 (0.0011)	-0.0391 (0.0006)
-	**Intersection density in terms of multi-modal intersections having three legs per square mile**	-0.0143 (0.0002)	0.0040 (0.0001)	-0.0078 (0.0001)	-0.0112 (0.0003)	-0.0010 (0.0001)

Base Category	Parameter	Wheelchair	Micro-mobility	Carpool	TNC or Taxi	Drove Alone
-	**Number of jobs per household**	0.0158 (0.0002)	-0.0040 (0.0001)	-0.0052 (0.0001)	-0.0156 (0.0003)	-0.0157 (0.0002)
-	**Regional Diversity***	0.1822 (0.0046)	-0.2039 (0.0026)	-0.1312 (0.0024)	-0.1877 (0.0076)	-0.1952 (0.0041)

* regional diversity measures the deviation of the CBG employment rate (jobs per person) from the regional average employment rate.

Table 16 Model Results for Egress Mode

Base Category	Parameter	Wheelchair	Micro-mobility	Carpool	TNC or Taxi	Drove Alone
	Intercept	-13.4066 (7.9537)	-4.7544 (0.3286)	-3.6420 (0.3349)	-6.7790 (0.7589)	-5.1356 (0.6185)
Egress Length	**Distance from a transit stop to the destination (mile)**		1.3488 (0.0785)	1.9382 (0.0730)	1.9516 (0.0743)	1.9496 (0.0738)
Destination Place	**Medical, Hospital**					1.6571 (0.5022)
	Shopping		-0.5440 (0.2947)	-0.8199 (0.4143)		
	University/college			-8.2498 (1.5404)		
Origin Place	**Airport**			1.6083 (0.3247)	1.2316 (0.6448)	
	Medical, Hospital	1.3308 (0.5081)				1.2256 (0.5121)
Transfer	**Number of transfers from the origin**		-0.2695 (0.1084)	-0.3085 (0.1304)		-1.7543 (0.3952)
	Number of transfers to the destination		-0.3313 (0.1084)	-0.3416 (0.1226)	-2.0377 (0.6283)	-1.3293 (0.3030)
Two-way trip	**Trip in the Opposite Direction-Yes**				-0.6058 (0.3155)	
Time Period	**Midday**			-0.5634 (0.1478)		-1.2591 (0.3002)
Month	**February**			-0.3214 (0.1499)	-0.8936 (0.4040)	
Disability	**Disability-Yes**			0.3710 (0.2085)		
Driver License	**Driver License-Yes**				0.9973 (0.3394)	2.7107 (0.4122)
Ethnicity	**African American**		-0.4305 (0.1279)	-0.5889 (0.1513)	-0.8613 (0.3531)	-1.7695 (0.2962)
	American Indian		0.8178 (0.4512)	-1.7185 (1.0373)		
	Asian				1.4318 (0.5830)	
	Hispanic		-0.3602 (0.1457)	-0.2862 (0.1597)		-1.1175 (0.2841)

Base Category	Parameter	Wheelchair	Micro-mobility	Carpool	TNC or Taxi	Drove Alone
Gender	**Male**					-0.4788 (0.2099)
HH Income	**Middle Income ($20K-$50K)**		0.5542 (0.1197)		0.9792 (0.3493)	
	High Income ($50K-$100K)		0.3693 (0.2035)	0.7323 (0.1930)	1.4262 (0.4377)	0.9137 (0.2805)
	Number of households in Destination CBG that own zero automobiles			-0.0016 (0.0007)		-0.0042 (0.0014)
	Gross entertainment (5-tier) employment density (jobs/acre) on unprotected land of destination			-0.0880 (0.0427)		-0.2173 (0.1125)
	Proportional Accessibility to Regional Destinations		3.9741 (1.3076)	3.8362 (1.1854)		4.8458 (1.2974)
	Regional Diversity of destination				1.4658 (0.4872)	0.7102 (0.3621)