SPATIOTEMPORAL PATTERNS AND INFLUENCING FACTORS OF SHARED BIKES USE FOR METRO ACCESS AND EGRESS TRIPS: A CASE STUDY IN XI'AN, CHINA

JIAXIN SUN
June, 2020

SUPERVISORS:
Dr. S. Amer
A. Da Silva Mano MSc
Ing.F.H.M. van den Bosch

SPATIOTEMPORAL PATTERNS AND INFLUENCING FACTORS OF SHARED BIKES USE FOR METRO ACCESS AND EGRESS TRIPS: A CASE STUDY IN XI'AN, CHINA

JIAXIN SUN

Enschede, The Netherlands, June, 2020

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Urban Planning and Management

SUPERVISORS:
Dr. S. Amer
A. Da Silva Mano MSc
Ing.F.H.M. van den Bosch

THESIS ASSESSMENT BOARD:
Dr. J.A. Martinez (Chair)]
Dr. T. Thomas (External Examiner, Transport Engineering and Management (TEM), University of Twente)

ABSTRACT

Dockless bike-sharing system (DLBS) as a potential mode to solve the “first and last mile” problem is commonly used as access and egress of metro systems in many Chinese cities. A better understanding of this novel, economical and environmentally friendly mode of public transport can help improve the efficiency of bike-sharing system, and enhance the connection between bike sharing and metro systems so as to bring convenience to the travellers. To do so, this study used different types of visualization techniques, based on the large-scale shared bike data, to explore the spatial and temporal use patterns of shared bikes use for metro system in Xi’an city, China. Besides, the geographically weighted regression (GWR) model was applied to reveal the relationship between the influencing landuse related factors and the use patterns of shared bikes.

Specifically, the Python-based data process method was used to deal with the large-scale GPS data of the shared bikes to extract useful information, such as the origins and destinations of the bike-metro trips. The cycling trajectory was also generated based on “Rout Plan” Application programming interface (API) from Baidu Map Server. A variety of statistical chart combinations, as well as different visualization methods such as heat maps, interactive flow maps, and routes distribution maps were used to extract the hot spots of the shared bike use. With the help of different visualization software or platforms, the spatial characteristics of the shared bike use in morning and evening peak hours were identified. In short, in the morning rush hours, there are a large number of shared bike trips that ride from the residential area to the metro station, or from the metro station to the job locations. While, in the evening peak hours, a large number of shared bike trips depart from the employment concentration to the metro station, or ride from the metro station to the residential area. This trend is consistent with the commuting activities on weekdays. Point of interest (POI) data was used as the basis for the influencing land use related factors The number of residence, job, commercial area, recreation area, green space, educational place, and health care locations were used as independent variables to explore their influence on the number of origins and destinations of bike-metro trips during different time slots. From the geographically weighted regression (GWR) model, in the morning peak, the number of residences has a positive effect on the distribution of the origins of the bike-metro trips, and the impact has spatial heterogeneity. The number of job locations and educational places has a positive influence on the distribution of destinations of bike-metro trips, and the influence has spatial heterogeneity. In the evening peak hours, the number of job locations has a positive impact on the distribution of the origins of bike-metro trips, and the influence has spatial heterogeneity. The number of recreation areas and residences has a positive effect on the distribution of the destinations of the shared bike use with spatial heterogeneity.

According to the results of this research, some suggestions were provided to improve the efficiency of the bike-sharing system so as to make shared bikes solve the "first mile/last mile" problem better.

Keywords: Dockless bike-sharing system, GPS data, Spatiotemporal characteristics, Visualization, GWR

ACKNOWLEDGEMENTS

With the deepest respect, I would like to express my sincerest gratitude to my supervisors, Dr Sherif Amer, Andre Da Silva Mano, and Frans van den Bosch for their wise and patient guidance, supervision throughout the process of thesis writing. Their patient consulting and valuable suggestions always encouraged me to accomplish this topic better.

I would like to thank all the staff at ITC for their help during my studies. And I am very grateful for having so many new friends during this time. I really appreciate the moments we shared our joy together through this program. Thank you for all their supports.

Special thanks to my peer of the research group from Chang'an University in China for providing me with support in research data. I really appreciate the moments we shared our life together through this program.

I would like to express my gratitude to the staff who are still working hard during the COVID-19 crisis to help us maintain a normal life in various fields.

Last but not least, many thanks to my family. Thanks to my mother and father for your love and support in all forms.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Background and Justification

Cycling as an environmentally friendly mode of transport, has been a growing interest for policymakers and academics, especially in highly motorized urban areas confronted with pollution and congestion (Canitez, 2019). In the field of transportation, one of the potentials for bicycles is to help solve the "first and last-mile" problem, which means, getting riders from their origin to access the boarding point of a public transport system and then egress from the destination transit station to final activities (Hall, 2012; de Souza, La Paix Puello, Brussel, Orrico, & van Maarseveen, 2017). A new concept called bicycle-transit integration has emerged. Bike-transit integration means that cycling is used as transfer mode to or from transit stations to achieve an efficient and sustainable urban public transport system (Zhao & Li, 2017; Krizek &Stonebraker, 2010). Governments have advocated the concept in both developed and developing countries such as Brazil, China, and Colombia (Shelat, Huisman, & van Oort, 2018; States, 2009). In short, cycling plays an important role in promoting the use of transit.

Among different public transit modes, some Asian countries rely on metro-based transport for the growing travel demand (Alam, 2010). The goal of metro construction is to improve sustainable transport, which can reduce negative environmental and social consequences caused by the rapid growth of motorization (Zhang & Zhuang, 2019). The efforts on improving the utilization rate of metro and enlarging the transit's catchment area by solving the first/last mile problem are ongoing in many Chinese cities (States, 2009; Wang, Chen, & Xu, 2016). With the characteristics discussed above, cycling has been an important feeder mode to expand the scope of metro service (Ma, Ji, Yang, Jin, & Tan, 2018).

The usage of bike sharing systems has become very popular in many Metropolis (Weliwitiya, 2019; de Souza, 2017; Shelat, 2018). Some research has revealed that bike sharing can significantly promote the integration of bikes and transit (DeMaio, 2009; DeMaio & Gifford, 2004). In 2008, the concept of docked-station based bike-sharing (DBS) was introduced in some Chinese metropolitan areas such as Beijing and Hangzhou. The municipal government leads the construction of DBS system as a part of the public transport system to cover short journeys and solve the first and last mile problem (Zhang, Duan, & Bryde, 2015). Users can rent a bike from public docked stations and return it at another station by using a transport smart card. However, because of the sparse distribution of docking stations, it can be quite difficult to reach the origins and the destinations of users' activities. Due to the flexibility limitation of station-based public bike-sharing, dock-less bike-sharing system (DLBS) are now becoming the most popular public bicycle system in China (Wang, Huang, & Dunford, 2019). This system provides on-demand service by mobile phone applications. The users can get a bike nearby by scanning its QR code and return it to any reasonable locations around the destination of a trip. Besides, this dockless system can save costs by avoiding the construction of expensive dock stations (Du, Deng, & Liao, 2019).

The cumulative registered users of shared bikes reached to 235 million in 2018. The overall estimated ridership of DLBS has reached 17 billion, with the highest peak use reaching 70 million rides a day in 2018. (Gu, Kim, & Currie, 2019; Du, Cheng, Li, & Yang, 2019; Ma, Shi, Yuen, Sun, & Guo, 2019). Besides, 90% of users choose to use shared bicycles when their destination is within 5000 meters from/to the metro or other transit stations (Mobike, 2017). Therefore, for some cities with the metro-led public transportation systems, the usage of shared bikes as access and egress mode for the metro system caused more and more concern.

With the large use of shared bikes, some researchers identify several problems caused by the new mode of shared bikes. For example, **parking disorder** caused by a large number of shared bikes is more and more serious; the **rebalancing** which refers to removing bikes from parking areas that near capacity to the areas that need bikes, is also a hot issue (Zhao, Ong, Wang, & Hu, 2019); the increase of the DLBS users makes the infrastructure of cycling unable to match in time, especially the **bike lanes**. Therefore, many relevant studies explored the spatiotemporal characteristics of the DLBS use to provide evidence for better dealing with these challenges and improve the efficiency of the system.

With the rise of GPS devices and smartphone-based APPs, more and more researchers realized that bike sharing system created an opportunity to analyse cycling activities on a vast data scale (Du, Deng, & Liao, 2019). Hence, data-driven methods are increasingly applied to analyse **spatiotemporal patterns** as a **basis** to help understand the use patterns of shared bikes.

Generally, the spatial pattern of bike sharing usage refers to the placement or arrangement of both **points** (available bikes' locations of the DLBS or the exact origins and destinations of the bike stations) and **routes**. Some researchers focus on the DLBS studying the availability of shared bikes showing all the locations of available bikes within fixed time intervals (Yang & Liu, 2018; Reynaud, 2018; Ashqar, 2017). Some research focuses on the distribution pattern of origins and destinations of cycling trips to examine the spatial pattern of bike use and its relationship to transit stations based on large volume GPS data with the information of users and bikes as well as the origin and destination coordinates (Chen, 2015; Faghih-Imani & Eluru, 2015; Wu, 2019; Barbour, 2019).

Meanwhile, some research on spatial patterns of shared bikes combines the data processing methods with **visualization tools**. For instance, Reiss and Bogenberger (2015) use "*heatmaps*" to show the rental-return pattern of DLBS in Munich. Yang, Heppenstall, Turner, and Comber (2019) use a travel "*flow map*" to show the origins and destinations distribution around the metro stations' catchment area in Nanchang. Cluster analysis was also used for spatial patterns of shared bikes usage combined with different land use, or Point of Interest (POI) data extracted from Application Program Interface (API) (Etienne & Latifa, 2014; Li, Zhu, & Guo, 2019). Besides, the pattern of cycling routes of bike-sharing is also a topic of some research based where trajectory data are used to suggest future bike lane construction (Yang, 2019; Barbour et al., 2019).

The temporal analysis of shared bike use is mainly about the usage variation at different moments in time. Kaltenbrunner, Meza, Grivolla, Codina, and Banchs (2010) provided a time series analysis to explore the temporal pattern of share bikes in Barcelona according to a 1-hour interval during about two months. Cluster analysis is an often used method for the exploration of temporal patterns of bike sharing usage. Zhang, Thomas, Brussel, and Van Maarseveen (2017) used clustering analysis to classify the bike activity patterns according to weekdays, weekends, peak-hour, and off-peak hours based on the DBS stations. In this way, the variation of activity patterns over time can be revealed, which can help to better understand the shared bike use patterns and manage the bikes efficiently. The similar clustering analysis for exploring the temporal activity pattern was shown in Vogel, Greiser, and Mattfeld, (2011), as well as Du's (2019) work.

Analysing influencing factors of shared bikes usage is a good way to better understand the spatiotemporal patterns. Zhao (2017) analyse factors from the built environment, transport infrastructure and service attributes. Factors like travel distance, travel time, income, car ownership, number of households, and number of public bikes etc. are considered. Lin (2018) and Xu (2019) focused on the impact of built environments. In Lin's study, factors were divided into five dimensions, which are density, diversity, design, distance to transit and distribution of shared bikes. For each dimension, for instance, population density, job-housing balance, street length, transfer distance and the number of shard bikes were included respectively. For Xu's study, several key factors were selected, such as floor area ratio (FAR) of residential buildings, FAR of commercial buildings, land use mixture, network distance to MRT (Mass Rapid Transit), number of bus stops, number of road intersects, and length of cycling lanes. Cole-Hunter (2015) analysed the relation of cycling commuting trips with environmental and socio-demographic indicators. Age, gender, and income are often used for socio-demographic dimension. Noise, elevation, greenness, bicycle lanes, bicycle racks, were used as environment factors for instance. Besides, weather conditions are concerned (Ashqar, 2019; Caulfield, 2017). Some related policies and the cost of using the bikes are also important to the usage of bike sharing (Griffin & Sener, 2016; Li, Zhang, Sun, & Liu, 2018).

1.2 Research Problem

Based on the justification, the literature exploring spatiotemporal patterns of the bikes use for different purposes is quite sufficient. Spatial patterns are mainly analysed from the distribution pattern of origin-destination or locations of available bikes as well as cycling routes pattern; temporal patterns are mainly identified according to different moments in time. Although different usage profiles of bikes were considered, a minority of the research focuses on the use patterns of bike sharing for the metro system (González, Melo-Riquelme, & de Grange, 2016; Faghih-Imani & Eluru, 2015). In addition, Some research discusses only the "bikes" used for connecting to public transit without distinguishing between private bikes and public bikes (de Souza, 2017; Wu & Yang, 2018). Furthermore, most of the research analyses usage patterns of shared bikes focus on the DBS system, which has different operation and management from the DLBS as mentioned before (Li, 2019).

As discussed previously, there are different variables used to explain the use pattern of shared bikes. Nevertheless, it is also worth noticing that the existing research on analysing influencing factors of bike sharing usage usually does not focus on the variables that influence the use of shared bikes for metro systems based on the vast volume of secondary data as opposed to primary survey data.

Based on these gaps, the study will develop an approach to analyse the spatiotemporal patterns and its influencing factors of DLBS used for connecting with the metro. This study aims to provide evidence for operators and urban planners so that they can better manage this new public transport mode, which can solve the "first/last mile" problem more efficiently.

1.3 Research objectives

1.3.1 General Objectives

The general objective of this research is to develop an approach to analyse the spatiotemporal patterns and identify the main factors influencing the usage of a dockless bike sharing system for metro using in Xi'an, China as a case study.

1.3.2 Specific Objectives

1. To identify the temporal patterns of dockless bike sharing use for connecting with the metro stations.

2. To provide insight in the spatial pattern of dockless bike sharing used as access and egress mode to metro stations.

3. To identify factors that influence the use patterns of bike-sharing for connecting with the metro stations.

1.4 Research Questions

1. To identify the temporal patterns of dockless bike sharing use for connecting with the metro stations.

- What are the suitable analytical methods to identify the temporal patterns of shared bikes usage for metro stations?
- Which method is most suitable to analyse the temporal patterns?

2. To provide insight in the spatial pattern of dockless bike sharing used as access and egress mode to metro stations.

- What methods can be used for identifying spatial patterns of bike sharing used as access and egress mode to metro system?
- Which method is most suitable to analyse the spatial patterns?
- What is the difference in the spatial pattern of bike sharing from and to metro stations?

3. To identify factors that influence the use patterns of bike-sharing for connecting with the metro stations.

- What are the factors that influence the usage of share bikes to metro stations based on literature?
- What analytical methods have been applied in studying the influencing factors of share bikes as access or egress of metro?
- Which method is most suitable for analysing the relationship between influencing factors and bike sharing use?
- What are the possible reasons for the spatiotemporal patterns according to the factors?

1.5 Thesis Structure

This study consists of six chapters:

Chapter 1 introduces the background and justification, research problem, research objectives and questions, as well as a conceptual framework.

Chapter 2 reviews the literature about the development of the DLBS in China, the research values and methods to identify the spatiotemporal characteristics of shared bike use, and the classification of influencing factors, as well as the techniques to deal with the factors.

Chapter 3 illustrates the study area, datasets and sources, as well as the method used to analyze the spatiotemporal characteristics of shared bike use and its influencing factors.

Chapter 4 visualizes and analyses the use patterns of shared bikes that connect with metro system, as well as the impact of the influencing factors. Besides, this chapter presents the discussion for the results, and some recommendations based on the results. A reflection on the datasets and methodology was also presented.

Chapter 5 presents the conclusions and recommendations for future research.

2 LITERATURE REVIEW

This chapter provides an overview of the related research about the spatiotemporal patterns of bike-sharing use and its influencing factors. The **first** part describes the current development of the shared bikes system in China, focusing on different purposes of using the system and the significance of shared bikes use for commuting trips. The **second** part mainly summarizes the general analytical methods used in different studies to identify the spatiotemporal distribution of shared bike use. In the **third** part, different visualization techniques are discussed based on different data types of shared bikes, so as to identify the spatiotemporal distribution characteristics of shared bike use. The **fourth** part summarizes the influencing factors that affect the distribution of shared bike use in different literature. This section also discusses the analytical methods used to analyse the relationship between these factors and the spatiotemporal distribution of shared bike use.

2.1 Bike-sharing in China

Due to its flexibility and convenience, the dockless bike-sharing system (DLBS), also known as the free-floating bike-sharing system, or the fourth generation of public bikes, has largely replaced the traditional bicycle sharing system in most metropolitan areas in China (Li, Zhu, & Guo, 2019).

This GPS-based novel public bike mode serves and facilitates many users with convenience and reasonable cost. Usually, for each bike, there is a unique QR code for users to unlock via a smartphone application. Users are required to register before the first ride and are charged with around 1 Yuan (around 0.2 Euro) if the trips are within half an hour or 15 minutes. The standard of charges is different among different operators (Ma, Zhang, Li, Wang, & Zhao, 2019).

Many studies focused on understanding the DLBS use for different purposes. From one of the largest dockless bike-sharing operators in China, the Mobike company, with the support of Tsinghua University and Beijing Tsinghua Tong Heng Planning and Design Institute, published a report to summarize the usage patterns in the cities that launched the DLBS bikes. The report revealed that the DLBS plus public transport is the most efficient way to get around, especially for trips shorter than 5km. Meanwhile, this integration also reduces the travel time of longer trips more than 5km. Besides, one-third of the shared bike users use bikes for leisure, while one-fifth integrate the bike-sharing ridership with public transport in China (Mobike, 2017). From these statistics, it can be seen that the development of the DLBS has great significance for commuters to solve the first and last mile problem. Moreover, in many metropolis in China, such as Beijing, most of the shared bike trips were found around bus and metro stations, about 81% and 44% respectively. In Shanghai, the numbers are 90% and 51%, respectively. (Chen, van Lierop, & Ettema, 2020; Mobike, 2017).

2.2 Spatiotemporal pattern identification of bike-sharing use

In line with the description above, many researchers find that the bike-sharing system is a good way to connect with other transport modes such as metro and buses to increase bicycle or public transport use and solve traveling the first and last mile problem (Wang, 2018). Based on the type of bike-sharing system, many studies focused on the station-based shared bikes as a feeder mode. Wu, Gu, Fan, and Cassidy (2020) compared the influence of station-based shared bikes on connecting to public transit and concluded that shared bikes are cost-effective as access and egress of public transit. Similarly, De Souza, La Paix Puello, Brussel, Orrico, and van Maarseveen (2017) focus on the integration of bicycles and bus, train as well as metro to identify the main factors that influence the bike use as a feeder mode to public transport. Besides, some studies explored the relationship between shared bikes and metro system. Gu, Kim, and Currie (2019) considered that these two modes should interact with each other rather than replace each other when the metro line density increased.

At present, there are fewer researches focused on the DLBS compared to the DBS, but more and more studies pay attention to better understand the DLBS. For example, Liu, Sun, Chen, and Ma (2019) demonstrated that dockless bike-sharing has an impact on the feeder bus of the metro system. From the study of Du and Cheng (2018), 51.3% of the DLBS users in the study area use the bikes to a transfer center of metro or bus, which is the most proportion comparing with the origin to destination trips and the travel cycle trips.

However, the widespread use of dockless shared bikes has also brought some **challenges**. To provide insight for the usage of this novel transport mode so as to better deal with these challenges, different analytical methods are used to identify the spatiotemporal characteristics of bike-sharing use. *Table2-1* summarized the related work on dealing with the challenges through spatiotemporal use patterns identification. Many studies discussed **parking disorder** as users can return the bikes to any reasonable locations without a fixed standard. *Cluster analysis* such as K-means clustering, density-based spatial clustering are useful to investigate the similarity of usage patterns for the origins and destinations of the bike trips, which can help to determine the potential locations of the designated parking areas to guide users to park the bikes (Vogel, Greiser, & Mattfeld, 2011; Li, 2019; Yan, Tao, Xu, Ren, & Lin, 2018; Lahoorpoor, Faroqi, Sadeghi-Niaraki, & Choi, 2019; Liu & Lin, 2019). The **rebalancing and reallocation** issue, which refers to moving bikes from oversupply areas to the areas that need bikes, is also a hot issue (Zhao, Ong, Wang, & Hu, 2019). The *matrix-based* methods are usually based on the origin-destination pairs of bike trips, which helps identify the variation of the departures/arrivals and gives an insight for **rebalancing issues.** Moreover, there are also some researchers pay attention to the **bike lane planning** to cope with the development of the DLBS. *Network-based* methods are often used to address the bike lane related issues from a spatial aspect. The technique concerns more about the topological relationship between nodes and links, which using points to represent the start and stop locations, and using lines or segments to reflect the paths of bike trips.

Table 2-1: Related work about analytical techniques to identify the spatiotemporal patterns of shared bike use

Elements	Techniques	Related work	Methods	Subjects/attribution	DBS/DLBS
Parking disorder	Clustering-based analysis	Zhang, Lin, and Mi (2019)	density-based spatial clustering	To identify the distribution of the origins and destinations of the trips so as to find out high density parking demand areas.	DLBS
		Du, Deng, and Liao (2019)	Hierarchical clustering algorithm	identified the temporal usage pattern of the DLBS in the different neighborhoods of different sites	DLBS
		Sun, Li, and Zuo (2019)	K-means clustering	To find out the characteristics of the locations of virtual stations of DLBS.	DLBS
Rebalancing	Matrix-based methods	Zhang (2018)	Transition probability matrix	To better understand the proportion of bikes in the bicycle parking area so as to further analyse the rebalancing process.	DLBS
	Network-based analysis	Pal and Zhang (2017)	Greedy network	To explore the spatial pattern of the shared bike use.	DLBS
Bike lane planning	Network-based analysis	Bao, He, Ruan, Li, and Zheng (2017)	Greedy network algorithm	To reveal the usage of the public bikes among different bike stations	DBS
		Yang, Heppenstall, Turner, and Comber (2019)	Graph-based visualization approach	To investigate the bike-sharing usage change after the introduction of a new metro line	DLBS

It is worth mentioning that more and more studies combining with visualization techniques to show the variations of the shared bike use from spatial and temporal aspects. This is a more intuitive way to give insight for understanding the shared bike use patterns from a more **comprehensive perspective**. Therefore, it can also give evidence to solve the management challenges mentioned above more comprehensively.

2.3 Visualization techniques for spatiotemporal identification of bike-sharing use

This section discusses the visualization methods according to different data contexts as well as visualization purposes. Generally, visualization methods enable researchers to better understand the time-varying pattern and spatial distribution of shared bikes use (Du, 2019; Oliveira, Sotomayor, Torchelsen, Silva, & Comba, 2016). The **first** part is about the data context used for identifying the spatiotemporal characteristics of shared bike usage, which are point datasets and trajectory data. The **second** part will discuss the possible visualization methods based on the point data. The **third** part will focus on the visualizations of trajectory data.

2.3.1 Data context for bike-sharing spatiotemporal patterns identification

The DLBS is equipped with GPS records and internet-controlled locks to provide online services for users. To some extent, it is an opportunity for transport operators and related researchers to obtain the automatically collected trip records related to renting, returning and real-time locations data of bikes (Du, Deng, & Liao, 2019)**.** Therefore, based on the large-scale data, some research pays attention to process the datasets and extract useful information to better understand the characteristics of shared bike use. In general, these related works can be classified into the visualizations based on **“point”** datasets, and the visualizations based on **“trajectory”** datasets.

On the one hand, due to the data availability, it is challenging to acquire the real trajectory data from open datasets. Thus, some researchers use **points datasets**. Unlike the station-based bike-sharing system, where the exact start and stop time and locations of trips can be obtained from the smart card, the point datasets of the DLBS used in most studies only consists of all the available bikes’ locations. The data scheme of this kind of dataset usually includes the timestamp, bike ID, and geolocation to show the available bikes’ distribution and provide information for users to locate the available bikes nearby. Some studies extract the origins and destinations of the bike trips from this type of dataset (McKenzie, 2019; Xu, Ji, & Liu, 2018; Wu, Wang, & Li, 2018; Lin, Zhang, Zhu, & Meng, 2019). There are also studies that combine the point dataset with other location-based datasets such as social network data to find out the use patterns of shared bikes with different spatial variables (Yang, Ding, Qu, & Ran, 2019).

On the other hand, some studies use **trajectory data** to analyse the spatiotemporal characteristics of bike-sharing use. Lin, Zhang, Zhu, and Meng (2019) used road network, metro station-related data, and dock-less bike trajectories from the Mobike company. The order ID, anonymous user ID, bike ID, and specific trajectory sampling points with timestamps and locations are included in the dataset. The authors modified the actual routes for shared bike trips by comparing them with the road network on the map based on the raw data for further identification. However, not all trajectory datasets can be used for analysis directly. Zhang and Mi (2018) estimated the trip distance of shared bikes to identify the spatial characteristics and users’ preference using a similar trajectory dataset, but the dataset used in the study lacks temporal information with only a collection of chronologically unordered track points.

Looking at the literature on the bike-sharing systems, whether DBS or DLBS, the spatiotemporal analysis of their use is basically based on the two types of data discussed here. However, minimal research generated the trajectories based on the point datasets. In order to have a more comprehensive understanding of the distribution of shared bike use, it is important to know the cycling routes hot spots for spatiotemporal identification.

2.3.2 Visualization methods for point data

Different visualization tools can help to illustrate the spatiotemporal distribution of shared bike usage. On the one hand, **line graphs, histograms, box plots** are often-used visualization methods to illustrate the temporal intensity or the distance distribution of shared bike usage. Yang, Heppenstall, Turner, and Comber (2019) used line graphs to show the daily variation of shared bikes usage in a week. They also extracted the origins and destinations from the available shared bikes dataset and made a histogram to show the association of the linear distance between the origins and destinations with the trip frequency. Some studies considered the travel time between the O-D pairs instead of linear distance to show users' time preference (Du, 2019; Wu, 2019). Besides, line graphs often used to indicate the variations of the groups generated by the clustering analysis. Li, Zhu, and Guo (2019) used *line graphs* (*Figure 2-1*) to reveal the differences among the five clusters generated by K-means clustering analysis based on the daily variations of the number of Mobikes near metro stations.

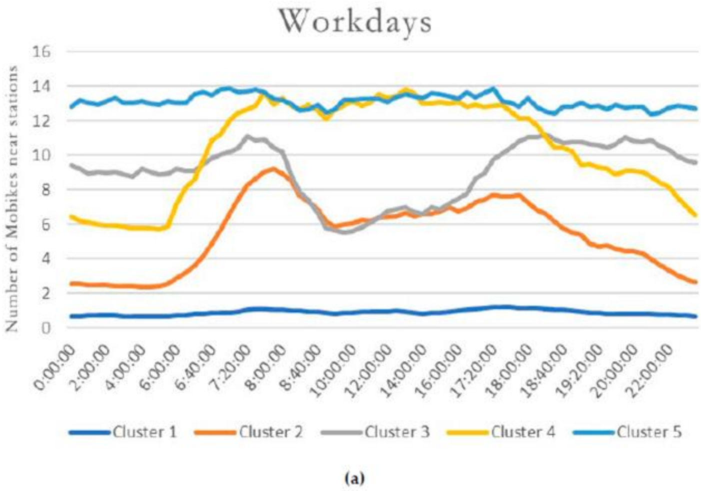


Figure 2-1: Temporal variation of clusters

(Source: Li, Zhu, and Guo (2019))

On the other hand, **heat maps or density maps** are used to show the hotspots of the shared bike use in the study area. In order to explore the relationship between the spatial distribution of bike trips and other traffic facilities or land use types, Du (2019) used a **heat map** (*Figure 2-2*) to show the hot spots of bike trips' origins and destinations based on different time slots, and they found the metro stations attracted most of the shared bike trips. **Density maps** are also used to identify the distribution patterns of trip origins and destinations. Yang (2019) made a Kernel Density map (*Figure 2-3*) in ArcGIS to reveal the spatial pattern variation of shared bike use around major metro stations.

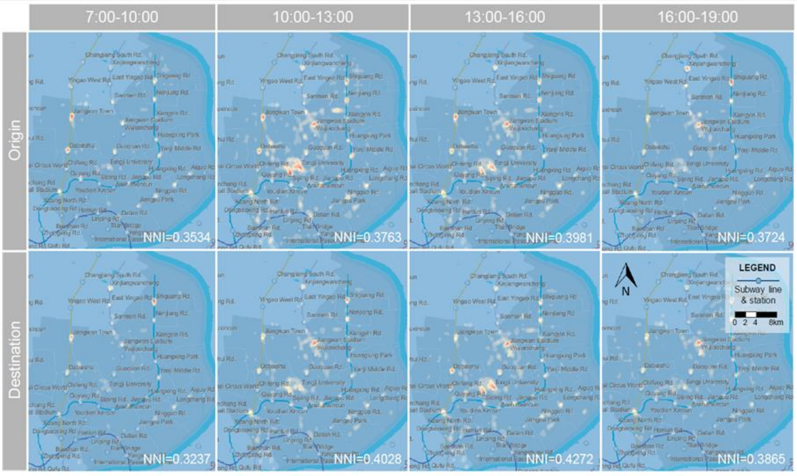


Figure 2-2: Heat map of origins and destinations distribution of shared bikes

(Source: Du (2019))

In addition, because the actual paths between the origins and destinations of the trip is uncertain according to the point dataset, **graph-based flow maps** (*Figure 2-4*) are used to show the number of trips between each O-D pair instead of studying the specific path. For example, Du (2019) and Yang(2019) used graph-based network analysis to make a flow map (*Figure 2-5*) with different thickness or schemed-colour lines represent the number of trips showing the information on the trip frequency around the metro stations. **Dendrogram** (*Figure 2-6*)is often used to show the temporal usage patterns of shared bikes around metro stations based on hierarchical clustering (Du, 2019). Moreover, **grid-based visualization** (*Figure 2-7*) methods are useful for matrix-based techniques such as the EigenDecomposition approach, trip chain, and transition matrices. The authors usually aggregate arrival trips or departure trips in a fixed-size grid cell to see the trip distribution characteristics (Zhang, 2018; Xu, 2019).

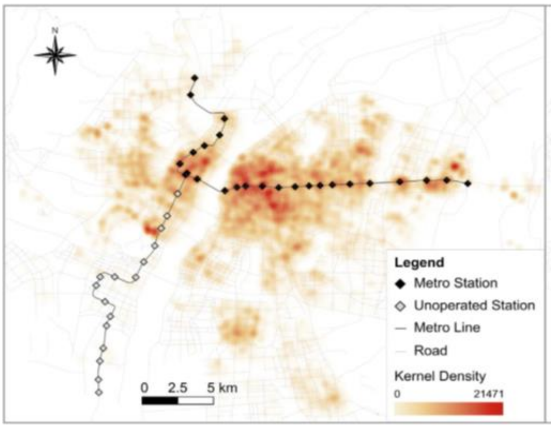


Figure 2-3: Kernel density of bike trips

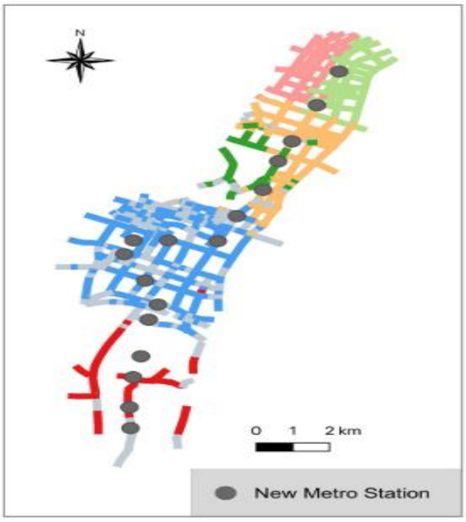


Figure 2-4: Road segments for DLBS network

(Source: Yang (2019))

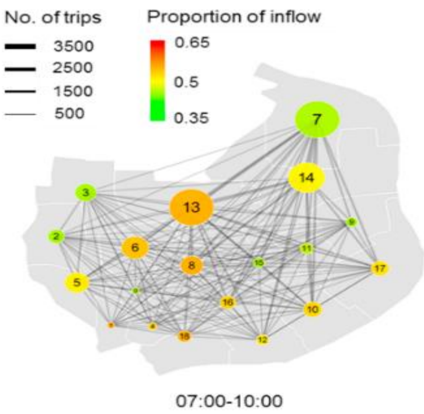


Figure 2-5: O-D proportion Flow map

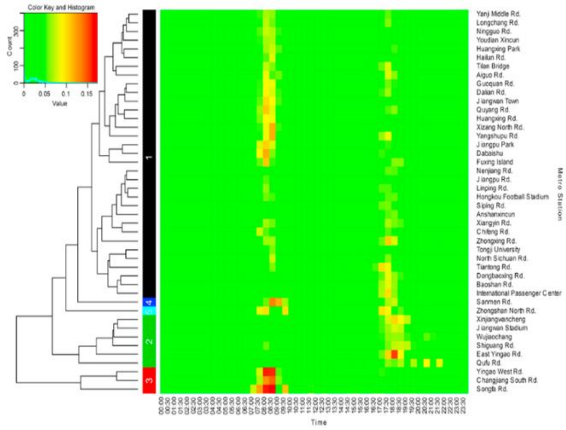


Figure 2-6: Dendrogram of Hierarchical Clustering results

(Source: Du (2019))

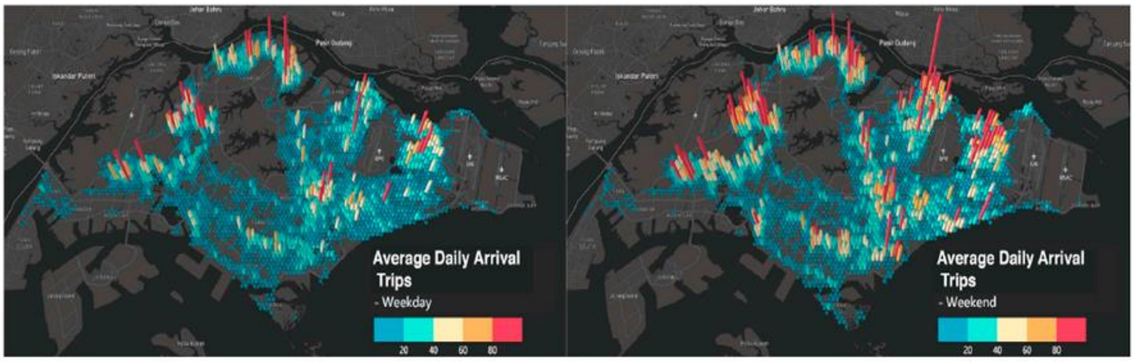


Figure 2-7: Grid-based visualization to show the daily bike trips (Source: Zhang (2018))

There are also some novel visualization techniques to combine more information into one figure. Du (2019) made a **space-time cube** (*Figure 2-8*) where x-y coordinate axes represent planar positions and the z-axis shows the time period to give an idea about the usage frequency of the DLBS combining both the spatial and temporal characteristics. Oliveira (2016) defined a new type of t**rips matrix** combined with geolocations (*Figure 2-9*) to reveal the incoming trips for selected bike stations located near a commuting hub on weekdays.

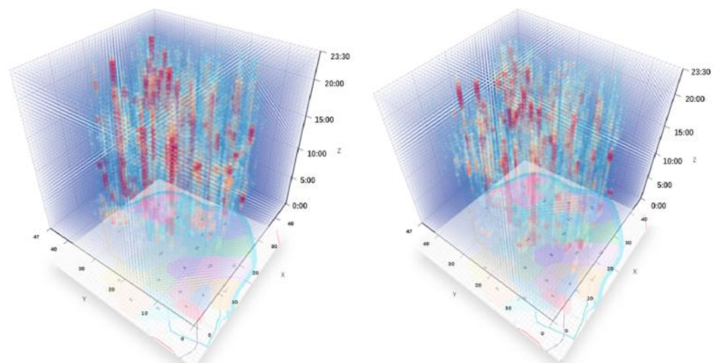


Figure 2-8: Space-time cube about check-in and check-out frequency

(a) (b)

Source: (Du, 2019)

2.3.3 Visualization methods for trajectory data

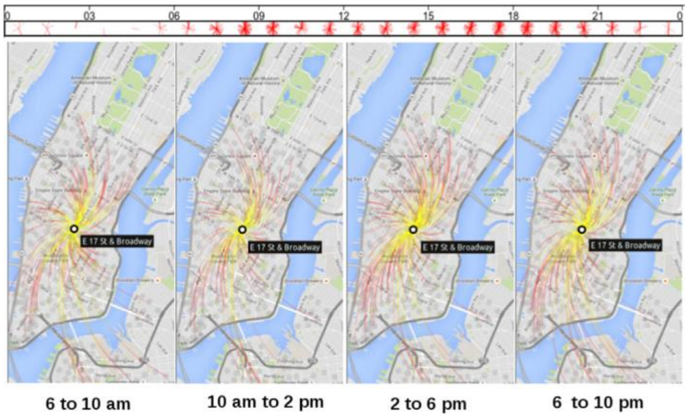


Figure 2-9: Trips matrix combined with geolocations

Source: (Oliveira, 2016)

Generally, visualizations based on trajectory data is not sufficient yet. Most of the studies that visualized this kind of dataset to show the most popular cycling routes. Bao, He, Ruan, Li, and Zheng (2017) used GPS trajectory data from shared bike company to construct a **heat map** for the paths of the trips to reveal the most frequently used paths as evidence for the bike lanes improvement. Besides heat maps, O'Brien (2014) estimated the bike trajectories based on the New York Citi Bike origin and destination data, and visualized the usage frequency of the routes with lines in different thicknesses, which is easy to identify the high utilization routes.

From the literature it can be concluded that, most of the visualization methods are used for the point data. The visualization methods used for trajectory data are very limited. In this research, different visualization techniques combined with big geodata processing will be used to identify the hotspots of the shared bike use patterns from both **O-D points and trajectory** aspects.

2.4 Factors associated with spatiotemporal characteristics of bike-sharing use

After analysing the spatiotemporal distribution of shared bike usage, many studies have further explored the factors that influence the spatiotemporal use patterns. This section **first** summarizes the factors associated with the bike-sharing use from different dimensions, and **then** summarizes the analytical methods used to reveal the relationship between these factors and the spatiotemporal characteristics of shared bike use.

2.4.1 Cost-related variables

The travel cost of the bike-sharing is very important for some users. Some studies also consider the cost of shared bikes as an influencing factor of using shared bikes. Li, Zhang, Sun, and Liu (2018) did a survey and considered the satisfaction of the DLBS fare as an indicator to show the user's attitude of the share bikes. Similarly, Guo, Zhou, Wu, and Li (2017) also did a survey and used satisfaction with the bike-sharing fee as a variable to affect the DBS bike usage. The studies revealed that satisfaction with the shared bike fee has a positive influence on bike usage. From these studies, cost-related variables are basically relevant to surveys. People with different income levels often have different degrees of satisfaction with the cost-related variable. Therefore, this factor is difficult to measure to some extent.

2.4.2 Socio-demographic variables

Some studies considered the influence of the users' personal characteristics, such as age, gender and employment type/status. Mobike (2017), indicates that 70% of the bike-sharing users are under 40 years old. Employees and college students have a high adoption of bike-sharing because they usually use the bikes for fixed commuting needs. From the report, there are no significant gender differences among users.

Most of the variables that come from the socio-demographic dimension are population-related. Du, Deng, and Liao (2019), used population size and proportion of migrants as indicators to predict the shared bikes use frequency in different districts of the case study area, Shanghai. Alcorn and Jiao (2019) analysed the surrounding population of the bike-sharing station in major cities in Texas to identify the relationship between bike-sharing utilization and population distribution. Similarly, Bao, Shi, and Zhang (2018) considered population density, non-white population as independent variables associated with public bikes ridership in the city of New York.

However, due to the availability of the data, the related information on bike-sharing users, such as age, employment is not easy to obtain from the secondary dataset (Lin, Zhang, Zhu, & Meng, 2019). Some of the studies did surveys for the information of users, which is very costly. But it is difficult to ensure the response rates and accuracy, which leads to a series of problems such as inadequate sample sizes and restricted study generalizability (Ji, Ma, Yang, Jin, & Gao, 2018). The population-related datasets are available, but often not sufficiently detailed at a neighbourhood or at a more specific level in case study area, which is not helpful for this research.

2.4.3 Built environment variables

Although many studies analysed the relationship between socio-demographic variables and shared bike trips. Most of the studies based on secondary GPS data, pay more attention to built environment variables. The definition of the built environment is the human-made environment that provides necessary settings for human activity (Roof & Oleru, 2008). In order to determine the factors affecting the use of shared bicycles and their spatial and temporal distribution, factors from the built environment are summarized into three main dimensions, which are **transport infrastructure, land use, and accessibility**.

For the dimension of transport infrastructure, the indicators related to the road conditions such as the road density, slope, road straightness, the number of road intersections, and the density of major or minor road are often considered; other indicators in this dimension focused on the infrastructure of specific transport mode, such as the density of bus stations, metro stations, and the public bike station, as well as the length of the bike-lane (Wu, Lu, Lin, & Yang, 2019; Ji, Ma, Yang, Jin, & Gao, 2018, Alcorn & Jiao 2019; Lin, Zhang, & Meng, 2019; Caggiani, Camporeale, Ottomanelli, & Szeto, 2018).

For the dimension of land use, three main sub-dimensions are used for analysis. First is the impact of **specific land use**, such as green area, commercial, or office area (Ma, 2019). Second, **land use diversity**, including Shannon entropy based on different categories of land use (Shen, 2018). Third, **land use density**, such as floor-ratio of residence or commercial buildings (Shen, 2018).

For the dimension of accessibility, two sub-dimensions are included. One is access to activities, such as work, leisure (Wu, Lu, Lin, & Yang, 2019); the other one is the distance to public transport, such as the distance to metro stations, bus stations, and the DLBS bikes (Zhao, 2019; Shen, 2018).

In summary, most of the related work paid more attention to the **land use** characteristics when they discuss about factors from built environment dimension, which is more physical and based mainly on secondary data. Thus, the research will not restrict by the sample size of the survey. Generally, this dimension contains almost all physical factors link to human activities, which is more related to the objectives of this study.

2.4.4 Weather-related variables

Weather-related variables are mainly about the influence of **temperature** and **precipitation** on the use of shared bikes. The studies considered this dimension are usually based on a *long-term* dataset to show the variation of weather for more than one week. For example, Shen (2018) analysed 10-day data of the dockless bike-sharing system in Singapore to understand the effect of accumulated precipitation and the average temperature on bike usage. Other research concerned the impact of weather can be found in the work of El-Assi, Salah Mahmoud, and Nurul Habib (2017) as well as Corcoran, Li, Rohde, Charles-Edwards, and Mateo-Babiano (2014). It can be seen from these studies that to investigate the relationship between the use of shared bikes and weather-related factors, it is necessary to have a long-term scope to ensure that the weather changes will affect the use of shared bikes.

2.4.5 Summary of the factors associated with the shared bike use patterns

All the indicators from the four main dimensions are summarized in the table below (*Table 2-2*).

Table 2-2: Summary of the related influencing factors

Categories	**Dimensions**	**Indicators**	**Reference**
Cost	Satisfaction	Satisfaction with the shared bike using fee	Li, Zhang, Sun, and Liu (2018); Guo, Zhou, Wu, and Li (2017)
Socio-demographic	Personal information	Age, gender, employment	Mobike (2017)
	Population	population and proportion of migrants, population density, non-white population	Du, Deng, and Liao (2019); Bao, Shi, and Zhang (2018)
Built-environment	Transport infrastructure	Road density, slope	Wu, Lu, Lin, and Yang (2019)
		Density of bus stations/metro stations/bike share stations/road network	Ji, Ma, Yang, Jin, and Gao (2018)
		Length of proximate high-comfort bikeways/sidewalk, primary road density	Alcorn and Jiao (2019)
		Road straightness, bus station number	Lin, Zhang, Zhu, Meng (2019)
		The number of road intersections	Caggiani, Camporeale, Ottomanelli, and Szeto (2018)
		Bike-lane length	Bao (2018); Du (2019); Shen, Zhang, and Zhao (2018)
	Land use	The number of hotel/ shopping mall/ living service/ leisure/ education/ real estate/ tourist site/ government/ corporations/ park and green area/commercial and office area/ residential area	Du (2019); Wu, Lu, Lin, and Yang (2019); Xinwei Ma, Cao, and Jin (2019); Ma (2019)
		Land use diversity	Wu, Wang, and Li (2018); Shen (2018)
		Land use density	Shen (2018)
	Accessibility	Access to Residence/work/commercial/park/leisure/ public transportation	Wu, Lu, Lin, and Yang (2019)
		Access to the dock-less shared bikes	Zhao (2019)
		Access to metro stations	Shen (2018)
Weather	temperature	Average temperature	Shen (2018); El-Assi, Salah Mahmoud, and Nurul Habib (2017); Corcoran, Li, Rohde, Charles-Edwards, and Mateo-Babiano (2014).
	Precipitation	Accumulated precipitation	

In conclusion, only the l**and use-related factors** from **built environment** variables are used to explore the relationship between shared bike use patterns and possible influencing factors in this research. The next section will discuss the promising methods to analyse the association of factors and shared bike use patterns.

2.4.6 Methods to reveal the relationship between the influencing factors and shared bike use

The related analytical methods can be classified mainly into two categories. First, some **regression methods** can help to explore the relationship between selected variables and the use frequency of shared bikes. Among them, geographically weighted regression is often used for analysing spatial heterogeneity because it allows users to estimate parameters at different places in the study area if the spatial coordinates are available (Ji et al., 2018). Second, there is also some related work focused on predicting the shared bike use based on the indicators from various dimensions through a more sophisticated approach such as Random Forest (RF). The related methods are listed in the table below (*Table 2-3*).

Table 2-3: Summary of Analytical techniques to analyse the influencing factors

Reference	Methods	subjects
Ma (2019); Yang (2019); Bao (2018)	Geographically Weighted Regression (GWR)	To quantify the influence of different spatial variables on bike-sharing usage.
Ji (2018)	Geographically Weighted Poisson Regression (GWPR) model	To investigate the relationships between the bikeshare-metro transfer frequency and built environment variables.
Liu and Lin (2019)	Multinomial logit regression (MNL)	To clarify the association of built environment variables with the spatiotemporal patterns of the public bike-sharing use.
Alcorn and Jiao (2019)	Stepwise multiple variable regression	To analyse the effect of different built environment indicators on the usage of the shared bike.
Du (2019)	Random forest	To evaluate the contribution degree of some potential variables that have an impact on bike use.
Qian, Pianura, and Comin (2018)	Boosting Gradient Regression (BGR)	To predict the number of rental and returns of bikes based on different built environment variables.

Generally, from the related research, different regression models are often used to explore the association between different variables and shared bike use patterns. Logistic regression is a very useful and straightforward method to model categorical outcome variables (Liu & Lin, 2019). However, the categorical variables in this method are binary decision variables (for example, Yes/No), which is not suitable for the shared bike datasets (Jakaitiene, 2018).

The global regression such as the linear regression that based on the Ordinary Least Square (OLS) model is often used to create equations that best describe the overall data relationships in a study area (Esri, n.d.-b). However, global model is hard to explain the regional variation. Thus, GWR model as a local model is commonly used as an extension of global model to estimate parameters locally and better portray the relationships among the dependent and independent variables (Zhao, Chow, Li, & Liu, 2005).

In general, geographically weighted regression (GWR) model is often proposed to address the spatial heterogeneity problem when modelling spatially aggregated data, which is suitable for the land use-related variables (Ma, 2019; Yang, 2019; Bao, 2018). This method helps users to estimate parameters at any place in the study area (Ji, 2018). Generally, the GWR model is widely applied in public transit forecasting and evaluating (Zhao, Chow, Li, & Liu, 2005), travel demand prediction (Cardozo, García-Palomares, & Gutiérrez, 2012), as well as traffic safety estimation (Bao, Liu, Yu, & Xu, 2017). It is easier for planners and transport engineers to understand and analyse traffic phenomena in practical application compared with the complicated machine learning method such as Random Forest (RF) or Boosting Gradient Regression (BGR) analysis, which are often used for prediction of the usage based on the predictor variables that have complex intersections without a distributional assumption about the variables (Bao, 2018; Du, 2019). **Thus, the GWR is the most promising method to analyse the influencing factors on the shared bike use in this study.**

2.5 Conceptual Framework

Based on the literature review, the conceptual framework (*Figure 2-10*) shows the main concepts in this study. The dockless bike sharing system has been a popular mode to connect with the metro as an environmentally-friendly way to solve the first and last mile problem in Chinese cities (Lin, Zhang, Zhu, & Meng, 2019; Chen, 2019). In order to promote the development of the DLBS system as an efficient way to connect to the metro, it is important to better understand the spatiotemporal patterns and the influencing factors of the DLBS activities. From literature, visualization techniques based on data processing are used to reveal the spatiotemporal characteristics of bike-metro trips. The influencing factors from different attributes were reviewed in the literature. The land use related factors from built environment dimension are determined to help explain the activity patterns of bike sharing usage through the geographically weighted regression.

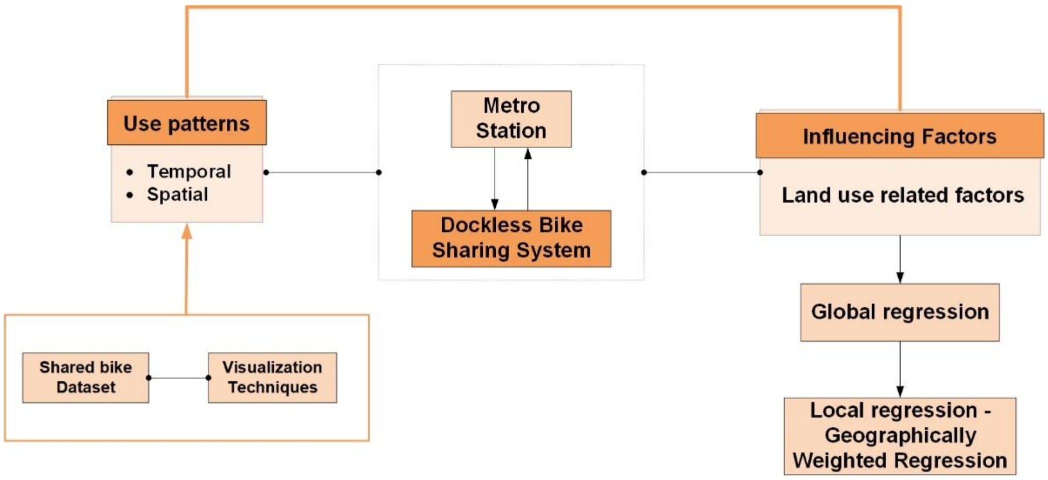


Figure 2-10: Conceptual framework

3 STUDY AREA AND RESEARCH METHODOLOGY

3.1 Study area

3.1.1 General description of study area

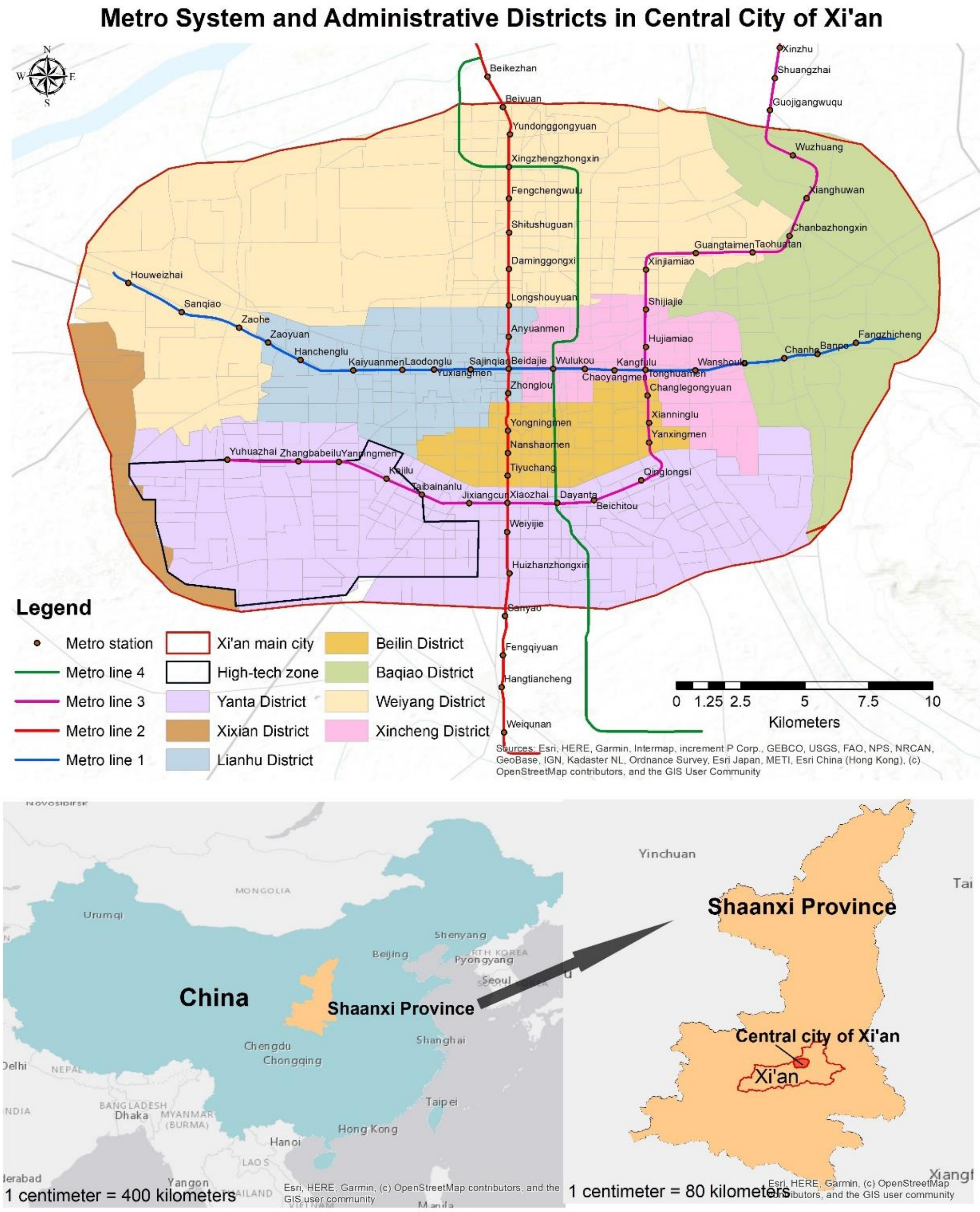


Figure 3-1：*Study area, Xi'an City, China*

In this study, a Chinese city, Xi'an (*Figure 3-1*), is selected as a case study city. It is the capital city of **Shaanxi Province** in the central of China. The total area of the city is 10,752 square kilometers, with a permanent population of 10.37 million in 2018. There are 11 districts and two counties under the jurisdiction of Xi'an Municipal Government, **7 districts** of which belong to the **central city of Xi'an**. The study area in this research only including the central city which locates within the city beltway (7 districts are included in the central city: District Xincheng, Beilin, Yanta, Lianhu, Weiyang, Baqiao and part of Xixian). There is a **high-tech development zone** located on the southwest of the city shows on the map.

There is a rapidly developing metro network in this city. Since the operation of the first metro line in 2011, by the end of 2018, 4 metro lines were operating in Xi'an, with a total mileage of 132.45 km and 99 stations. The average daily traffic volume of metro system in Xi'an reached 2,020,000 passengers, ranking ninth in China in 2018 (Official website of Xi'an Metro, 2019). Xi'an is also one of the cities that allowed the launch of bike sharing in the early stage of the DLBS development. There were multiple operators of shared bicycles in Xi'an by the end of 2017. However, the boom of the DLBS has also brought some challenges for management (Chen, 2019). Therefore, regulations have been enacted to control the number of shared bicycles put into the market by the Ministry of Transport (Ministry of Transport of the People's Republic of China, 2018). Recently, only three different operators were agreed to provide shared bikes into operation in the city, so the number of shared bicycles tends to be stable. By March of 2019, more than 450,000 bicycles have been put into the market by the three enterprises that have obtained the admission qualification (Xi'an Municipal Government, 2019).

3.1.2 The development of shared bikes in Xi'an

First, by the end of the first half of 2017, a bike-sharing company called "ofo" has launched about 100,000 shared bicycles in the market of Xi'an. Then the Kuqi and MoBike companies successively landed in Xi'an, with the market launch of about 90,000 and 100,000 (IXi'an, n.d.). After that, there are more and more brands of bike-sharing came into the market, which resulted in a blowout in the number of bicycles. Therefore, by the end of 2017, the office of Xi'an traffic management committee, in conjunction with four departments including the Municipal Urban Administration Bureau, the Municipal Administration for Industry and Commerce, and the traffic police detachment of the Municipal Public Security Bureau, jointly notified four major DLBS operation enterprises (ofo, Kuqi, MoBike and oxo) and asked all bike-sharing enterprises to "suspend the release of bikes" (Xinhua News, 2018). Due to the competition among different DLBS companies and government regulation, as well as some financial issues, some bike-sharing operators have declared bankruptcy and gradually withdrew from the market. As the development of bike-sharing in Xi'an, the bikes from Kuqi and oxo have disappeared, and the company of ofo has also declared bankruptcy. Another two brands were allowed to enter the market and added 90,000 operating bikes. So far, the number of shared bikes in Xi'an has reached 510,000, which including the bikes mainly from MoBike, HelloBike, and Qingju bike. This novel public transport mode serves and facilitates many users.

3.2 Research methodology

In this section, all the methods used in this study will be introduced. **First**, a work flow of all the methods will be summarized. **Second**, the various datasets and the sources used for the research will be described in detail. **Third**, based on the data used, the following section will discuss about all the data processing procedures to extract useful information for further research. After completing the data processing, the next phase is to identify the spatiotemporal distribution of the shared bikes use in the study area using different visualization techniques. The hot spots of the origins and destinations, and cycling routes will be identified. **Finally**, the factors related to land use will be analysed with the observed cycling use patterns to provide evidence for the improvement of the shared bike system.

3.2.1 Research approach

In general, this research is mainly based on large-scale data processing and visualization based on the GPS point data of shared bikes. Specifically, the methodology of the research can be divided into three blocks (*Figure 3-2* below). The **first** block is data processing in the environment of Python 3.4 to extract origin and destination information from the shared bike dataset, which forms the basis for the subsequent analytical steps. The **second** block analyses the spatiotemporal distribution characteristics of shared bikes using the processed data with origin and destination information. The **third** block will focus on the influence of factors related to the land use on the characteristics of shared bike use based on the riding origin-destination distribution. The following flowchart summarizes the design of the methodology for this research.

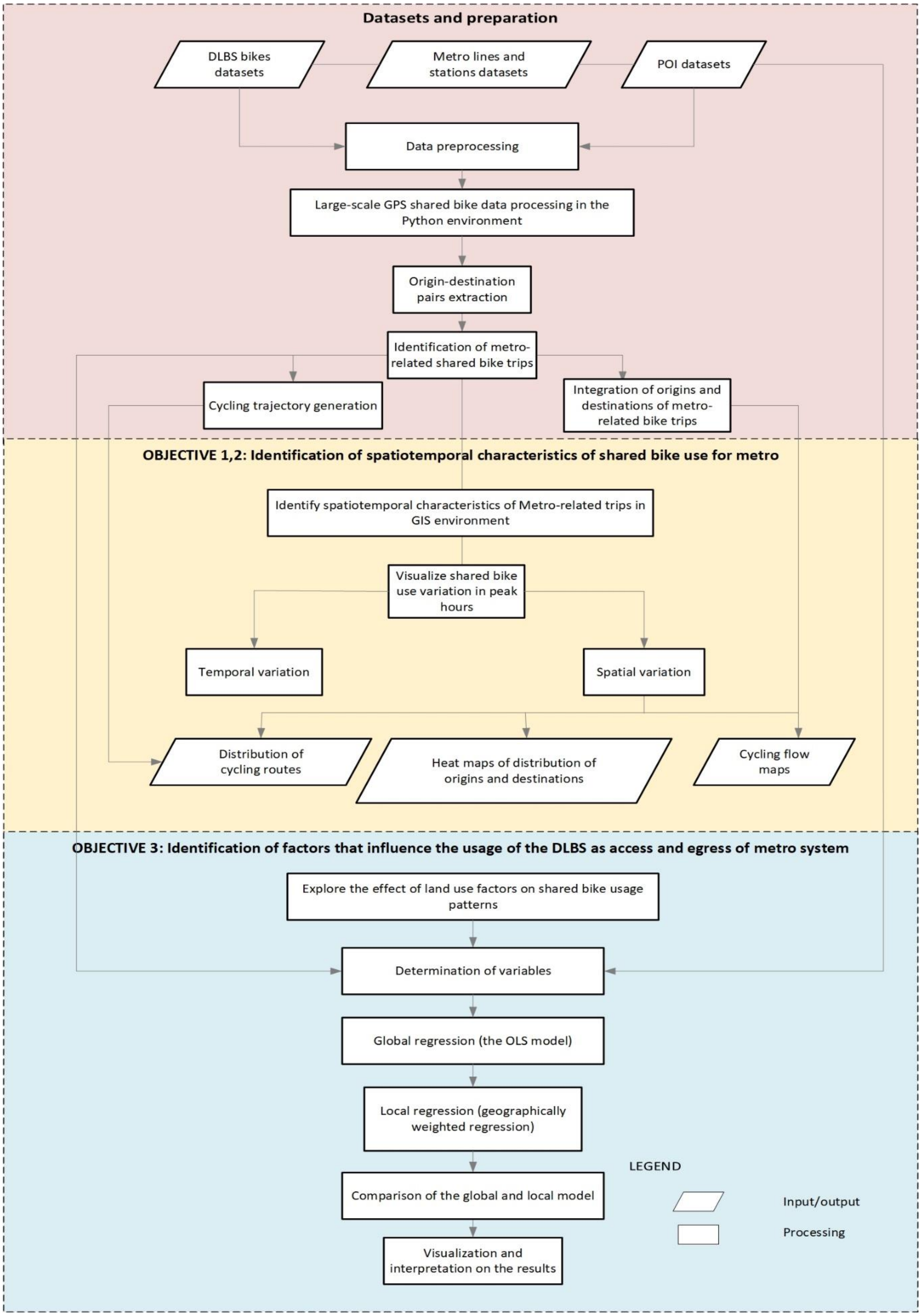


Figure 3-2: Methodology work flow

3.2.2 Datasets and sources

Based on the research design, this section will introduce all the datasets used in the study. All the data used in the research is secondary data. The datasets include: 1) the 3 workdays shared bike data which shows the available bikes; 2) the Point of Interest data extracted from a Chinese Map Server, Baidu Map (Baidu, 2020); 3) transportation data includes the road network and the metro system.

3.2.2.1 Bike Sharing GPS data

As the focus of the research is to analyse the spatiotemporal patterns, stated data from survey is not able to ensure randomness and sufficient samples. Thus, program was set up to collect DLBS availability data via application programming interface (API). The data is from a bike sharing company with the largest market share in Xi'an, named Mobike. The metadata consists of available bikes with the temporal resolution for every minute. The data can only be collected when the bikes keep still and once the bikes moved, the GPS data would not return to the system. Considering the weather conditions and the data availability, three workdays data without strong wind and rainfall was included, which are September, 3rd (Monday), September 7th (Friday), and September 10th (Monday) in 2018. The three-day data stored in three different files, which contains about 600 csv files respectively. For each csv file, the average number of the records is 150,000. Thus, the total number of the daily records in the raw data is around 90 million. The data frame including the time stamp, ID of the locked bike with longitude and latitude. The format of data is illustrated in *Table 3-1* below.

Table 3-1: Example of bike sharing data available

Timestamp	Bike ID	Longitude	Latitude
2018/9/3 8:39	8610034043#	108.8678	34.19164
2018/9/3 8:39	0296575157#	108.8679	34.19168
2018/9/7 8:40	8630826701#	108.617	34.30878
2018/9/10 8:40	0296541132#	108.6085	34.30835

3.2.2.2 Points of Interest (POI)

The points of interest data were often used as land use in the related work. POI is defined as a specific point of location that someone may find useful or interesting. From the study of Faghih-Imani, Hampshire, Marla, and Eluru (2017), the Points of Interest (POI) can result in higher trip generations and attractions. Most consumers use this term when referring to restaurants, hotels, park or any other categories applied in digital maps. Some previous studies have suggested that the POIs associated with different public transport systems have great potential to reveal travel patterns and possible trip purposes (Bao, Shi, & Zhang, 2018). Therefore, POI as a new type of electronic map data of land use can quantify the precise land use and related infrastructure characteristics, which provides a good way for many research to reveal the influence of different land-use types on a bicycle sharing system (Ma, 2019).

Many studies investigated the relationship between the land use and the shared bike use patterns by POIs data instead of the traditional land use dataset. For example, Zhao et al. (2019) and Bao (2017) used the POI data to explore the relationship between the hot spot distribution of cycle routes and land use type to provide a reference for bike lane construction. There are also some other researchers combined the POIs data with the bike parking hot spots to estimate the shared bike use patterns (Yang, Ding, Qu, & Ran, 2019; Shen, Zhang, & Zhao, 2018). From these studies, the POIs can help explain the distribution of origins and destinations of shared bike trips, and the number of a specific type of POI (e.g., residences, employments) as a factor is useful to analyse the influencing factors of bike sharing usage. Thus, in this research, the POIs dataset is used as the source of land use related factors for the analysis.

- **POI data extraction**

Datasets of POIs were obtained from a Chinese web mapping service application called Baidu Map using an application programming interface (API). The process of extracting the Points of Interest datasets including several main sub-processes. **First**, register an account to get a key for developer certification on the open platform of Baidu Map. **Second**, obtain the coordinates of the city boundary, and calculate the maximum and minimum longitude, latitude to cover the city by a large rectangular range. Third, according to a certain distance, divide the large rectangle into some small rectangles, and obtain the boundary coordinates of the small rectangles. **Then**, require the POIs of Baidu map, import the coordinates of the small rectangle borders, and the name of different POI types to get the data. The required POI datasets based on the Baidu coordinate system cannot be visualized directly, thus, a coordinate system transformation process is necessary. The flowchart in the *Appendix 1*shows all the processes to acquire the Points of Interest datasets.

- **Example of extracted POI data**

The data consists of many different categories of facilities: residence, job, commercial, life service, green space, recreation, education, health care, government (See the official classification in the *Appendix 2*). The information of all the classes is similar with the table *(Table 3-2)* below.

Table 3-2: An example of residence POI data

Name	Longitude	Latitude	District	Address	Type
Changle Fang Community	108.9706	34.2646	Beilin District	Wudao Shizi South Street	residence
Jiayi Community	108.9409	34.2279	Yanta District	104, Chang'an Road	residence

3.2.2.3 Road Network Dataset

The land use and road network dataset in Xi'an city can be obtained from Open Street Map. As the road network dataset is also from the year 2018, which the metro line 4 was still not opened to the public. Thus, the figure below only showed the location of the line 4 without stations (*Figure 3-3)*.

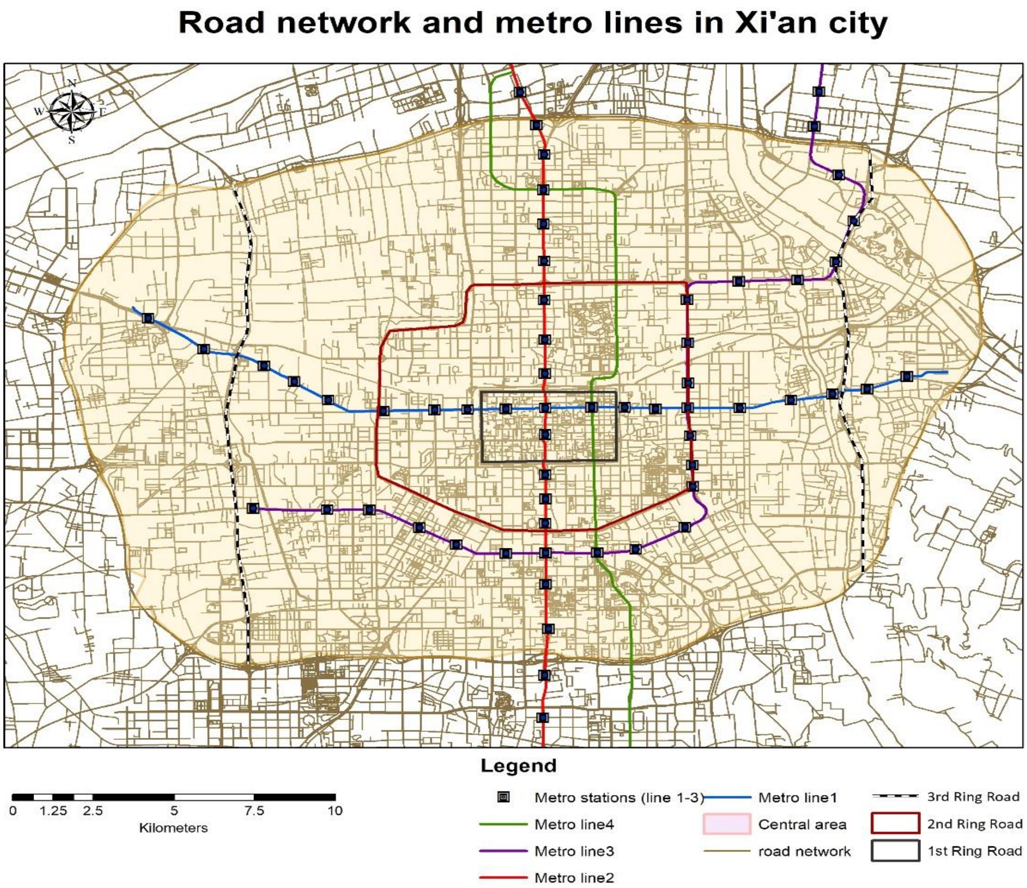


Figure 3-3: Road network and distribution of metro stations (Line1-3)

3.2.2.4 Data summary

Table 3-3: Data summary

Dataset	Format	Date	Source
Mobike GPS data	csv files	September3,7,10 (2018)	Mobike API
Points of interest	csv files	2018	Baidu Map Server
Road network data	shapefile	2018	Open Street Map

3.2.3 Data pre-processing

(1) Duplicate data cleaning

According to related research, when the shared bikes are locked, it is possible to see the GPS abnormality, and multiple records are returned at the same time. In this case, all the duplicated records need to be removed. Therefore, the three-day shared bike data is checked according to the bike id. If there are multiple pieces of returned records and the field information is consistent, the duplicate data needs to be deleted.

(2) Define the geographical boundary of data

The shared bike and POI data contain a large amount of data outside the study area, which will increase the computation time. Thus, filter all these data. It is necessary to determine the latitude and longitude boundaries according to the delineated study area. Therefore, the latitude and longitude boundary of the study area is identified by the coordinate identifier of Baidu map. In the Python environment, enter the range of longitude and latitude, and filter the data according to Equations in *Appendix 3*, leaving only the shared bike data located in the study area.

(3) Filter outliers

The shared bike data for each day has cross-day records that start at midnight and continues until the next day. These records are not into consideration in this research. The filtering method is: the date and time of the timestamp field are extracted using "*datetime*" function in the Python environment. Filter the records with cross-date.

3.2.4 Data processing

The data processing procedures described in this section are the basis for the following visualizations. The extraction of the O-D of shared bike trips are the most important step. The other processing procedures are all based on this.

3.2.4.1 Extract origin-destination trips from the available shared bikes dataset

Extracting useful origins and destinations information of the shared bike trips from the raw data is mainly performed in the *Python 3.4* environment, which can be mainly divided into five processes.

1) **Step 1**, the raw data needs to be integrated on a daily basis, and all records in the csv files contained in the folder of each day are organized into a new complete data frame.

2) **Step 2**, in the new data frame generated in the previous step, sort the values by the "id" and "time" columns of the table; A first-order difference function is used to compare the difference between the longitude or latitude record of the same bike after the time order and its previous as well as next records. Diff(-1) represented the longitude value minus the longitude value of the previous row, diff(1) means the longitude value of this row minus the value of next row. Filled all the missing values as zero. All the results of that are not zero from the difference function are reserved and stored into the new data frame. *(Figure 3-4)*

time	id	longitude	latitude	on	off
2018-09-10 01:46:00	0106600976#	108.948656	34.185375	-3.379248e-04	0.000000e+00
2018-09-10 01:48:43	0106600976#	108.948994	34.185445	0.000000e+00	3.379248e-04
2018-09-10 05:47:11	0106600976#	108.948994	34.185445	-2.799624e-05	0.000000e+00
2018-09-10 05:49:46	0106600976#	108.949022	34.185465	0.000000e+00	2.799624e-05
2018-09-10 09:46:45	0106600976#	108.949022	34.185465	-1.599368e-05	0.000000e+00
...	...	...	...	...	...
2018-09-10 01:20:56	9310002526#	108.742660	34.357812	0.000000e+00	-4.400323e-05
2018-09-10 05:15:41	9310002526#	108.742660	34.357812	4.744720e-09	0.000000e+00
2018-09-10 05:18:26	9310002526#	108.742660	34.357748	0.000000e+00	-4.744720e-09
2018-09-10 09:16:11	9310002526#	108.742660	34.357748	-5.101576e-05	0.000000e+00
2018-09-10 09:18:24	9310002526#	108.742711	34.357786	0.000000e+00	5.101576e-05

3229495 rows × 5 columns

Figure 3-4: An example of the result of "diff" function

3) **Step 3**, copy the data frame obtained in the previous step and delete the first record in the left table and the last record in the right table followed by joining the two tables horizontally. Rename the header of the table to get a new data frame containing the origin-destination information. *(Figure 3-5)*

start_time	id	start_x	start_y	stop_time	stop_x	stop_y
2018-09-10 01:46:00	0106600976#	108.948656	34.185375	2018-09-10 01:48:43	108.948994	34.185445
2018-09-10 05:47:11	0106600976#	108.948994	34.185445	2018-09-10 05:49:46	108.949022	34.185465
2018-09-10 09:46:45	0106600976#	108.949022	34.185465	2018-09-10 09:49:06	108.949038	34.185510
2018-09-10 13:46:02	0106600976#	108.949038	34.185510	2018-09-10 13:48:19	108.949046	34.185490
2018-09-10 17:45:34	0106600976#	108.949046	34.185490	2018-09-10 17:47:22	108.949019	34.185576
...	...	...	...	...	...	...
2018-09-10 13:25:17	8716025341#	108.969946	34.301600	2018-09-10 13:27:10	108.970060	34.301553
2018-09-10 17:26:12	8716025341#	108.970060	34.301553	2018-09-10 17:27:46	108.970113	34.301541
2018-09-10 21:25:18	8716025341#	108.970113	34.301541	2018-09-10 21:27:05	108.969971	34.301531
2018-09-10 01:16:54	9310002526#	108.742704	34.357811	2018-09-10 01:20:56	108.742660	34.357812
2018-09-10 05:15:41	9310002526#	108.742660	34.357812	2018-09-10 05:18:26	108.742660	34.357748

Figure 3-5: Results of new O-D dataframe

4) **Step 4**, sort out the "time" of the origins and destinations, and extract the time from the datetime to calculate the duration between OD pairs (unit: Second). The output table shows in *Figure 3-6*.

5) **Step 5**, using mathematical method to

calculate Linear Distance between each O-D pair according to the longitude and latitude values. The calculation formula is in *Appendix 4*.

After the Origin-destination extraction, several bikes were selected to plot all the GPS locations for three days to check the accuracy of the GPS service. The locations of the same bike changed a very short distance during a very short time period are considered as the GPS error. Thus, the trips with linear distance less than 100 meters or the duration less than 1 minute were removed to avoid the GPS deviation.

start_time	id	start_x	start_y	stop_time	stop_x	stop_y	STARTTIME	STOPTIME
2018-09-10 01:46:00	0106600976#	108.948656	34.185375	2018-09-10 01:48:43	108.948994	34.185445	01:46:00	01:48:43
2018-09-10 05:47:11	0106600976#	108.948994	34.185445	2018-09-10 05:49:46	108.949022	34.185465	05:47:11	05:49:46
2018-09-10 09:46:45	0106600976#	108.949022	34.185465	2018-09-10 09:49:06	108.949038	34.185510	09:46:45	09:49:06
2018-09-10 13:46:02	0106600976#	108.949038	34.185510	2018-09-10 13:48:19	108.949046	34.185490	13:46:02	13:48:19
2018-09-10 17:45:34	0106600976#	108.949046	34.185490	2018-09-10 17:47:22	108.949019	34.185576	17:45:34	17:47:22
...	...	...	...	...	...	...	...	...
2018-09-10 13:25:17	8716025341#	108.969946	34.301600	2018-09-10 13:27:10	108.970060	34.301553	13:25:17	13:27:10
2018-09-10 17:26:12	8716025341#	108.970060	34.301553	2018-09-10 17:27:46	108.970113	34.301541	17:26:12	17:27:46
2018-09-10 21:25:18	8716025341#	108.970113	34.301541	2018-09-10 21:27:05	108.969971	34.301531	21:25:18	21:27:05
2018-09-10 01:16:54	9310002526#	108.742704	34.357811	2018-09-10 01:20:56	108.742660	34.357812	01:16:54	01:20:56
2018-09-10 05:15:41	9310002526#	108.742660	34.357812	2018-09-10 05:18:26	108.742660	34.357748	05:15:41	05:18:26

Figure 3-6: Result of extracting start and stop time

Considering the objective of this research is the commuting trips used to connect with metro system, thus, the trips with linear distance more than **5km** or the O-D duration more than **1 hour** were also not considered. The final output table with the linear distance and duration shows as *Figure 3-7*.

	start_time	id	start_x	start_y	stop_time	stop_x	stop_y	STARTTIME	STOPTIME	Duration	Distance
0	2018/9/10	0290059856#	108.858326	34.272752	2018/9/10	108.856417	34.280777	00:01:28	00:18:54	1046	909.33
1	2018/9/10	0296537315#	108.870133	34.274464	2018/9/10	108.870570	34.288076	00:01:36	00:12:22	646	1514.11
2	2018/9/10	8630256015#	108.903094	34.224257	2018/9/10	108.927023	34.223981	00:01:45	00:16:38	893	2200.26
3	2018/9/10	8630240962#	108.912121	34.225994	2018/9/10	108.909852	34.209962	00:01:50	00:25:46	1436	1794.80
4	2018/9/10	0290019999#	108.929518	34.223770	2018/9/10	108.930772	34.224713	00:02:00	00:05:52	232	155.81
...	...	...	...	...	...	...	...	...	...	...	...
103668	2018/9/10	8630972721#	109.025594	34.312424	2018/9/10	108.990302	34.278975	23:53:36	23:56:58	202	4934.00
103669	2018/9/10	8630970001#	109.024954	34.312442	2018/9/10	108.991075	34.279543	23:53:36	23:56:59	203	4802.89
103670	2018/9/10	8630934468#	109.024448	34.312456	2018/9/10	108.990545	34.279737	23:53:36	23:56:59	203	4789.14
103671	2018/9/10	8617031691#	109.024255	34.309716	2018/9/10	108.990063	34.278824	23:53:36	23:56:58	202	4654.67
103672	2018/9/10	8630686879#	109.024283	34.309438	2018/9/10	108.990600	34.279287	23:53:36	23:56:59	203	4562.30

Figure 3-7: The final results

All the workflow of origin-destination extraction shows in the flow chart in *Figure 3-8*.

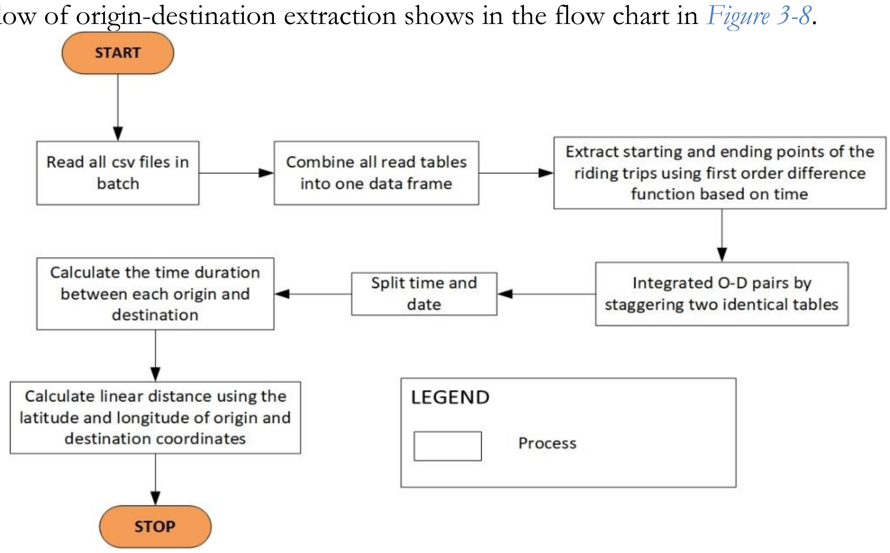


Figure 3-8: The final results

3.2.4.2 Identification of metro-related bike trips

In this study, an assumption is made to identify the DLBS-metro trips. The **assumption** is that, the origin or the destination of a bike-metro trip is located in a **100-meter** buffer zone of a metro station entrance. This assumption is consistent with that of Wu (2019). To identify the related trips based on the assumption, a filtering process was operated in *ArcGIS 10.7*. As the imported dataset in the ArcGIS included both the origin and destination information in the same row, the origin or the destination of each row can only be displayed in the layer at a time. Thus, the process to identify the metro-related trips need to consider both the origins and destinations of trips, which means the process need to import data with the origins (start point) and destinations (stop point) respectively. Besides, there are usually more than one entrances for each metro station in Xi'an city, all the detailed geolocation of the entrances was obtained from the "entrance and exit" category of POIs data. The "spatial join" function was used in ArcGIS to distinguish the origins (start points) or the destinations (stop points) that located within the buffer zone. The input "Target Feature" is the point layer, and the "Join Feature" is the buffer polygon layer. The value of the "Join_count" field in the attribute table generated after the spatial join operation can be used to determine the point in the buffer area. *Figure 3-9* simply shows the origins and destinations of bike-metro trips located within the buffer area. *Figure 3-10* is the flow chart of the whole process of metro-related shared bike trips identification.

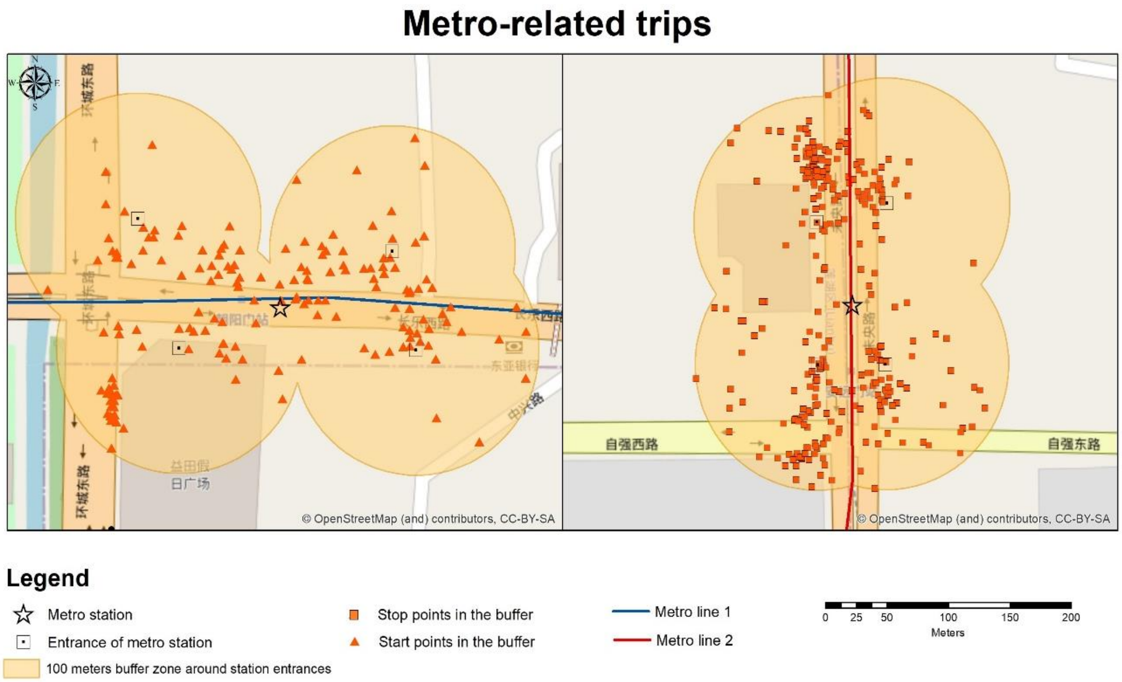


Figure 3-9: Example of the process to identify metro-related trips

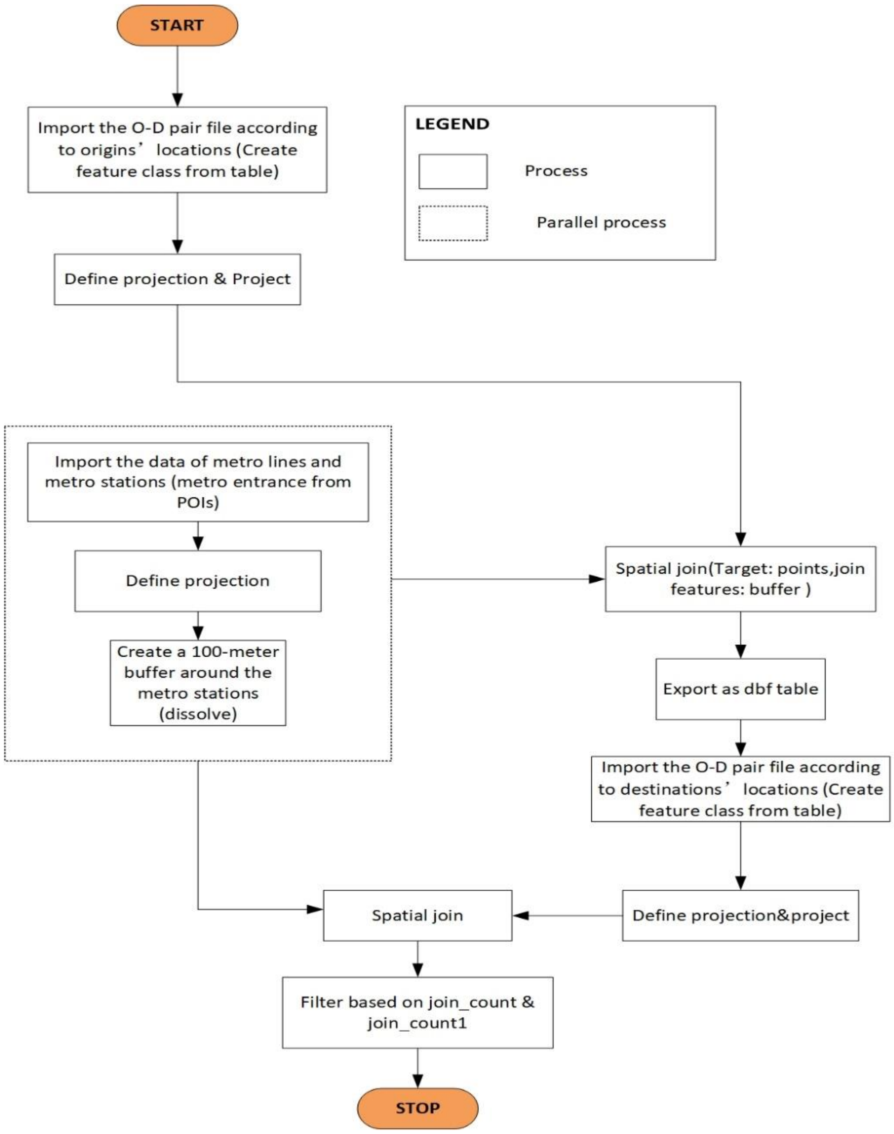


Figure 3-10: Work flow of identifying metro-related bike trips

3.2.4.3 Integration of origins and destinations of metro-related bike trips

Specifically, after identifying the shared bike trips that connect with metro stations, two different types of metro-related trips are defined here: the origin points locate in the buffer area of the metro station entrances and the destination points locate in the buffer of the metro station entrances. As the distribution of the origins of "toward metro station" trips and the destinations of the "from metro station" trips are scattered, it is difficult to find out a trend of the cycling flows. Therefore, a method was provided to integrate the origins and destinations onto the **centroids of 100*100 m² cells**. Specifically, "spatial join" function in *ArcMap 10.7* was used to generate the information about the ID of the metro stations (as origins) and destinations in the "from metro station" trips, and the origins and metro stations (as destinations) in the "toward metro station trips". To obtain the number of trips between each integrated O-D, a calculation based on "groupby" operation of "pandas" package in *Python 3.4* environment was used to calculate the real trips count from the metro stations (as aggregated origins) to aggregated destinations, or from the aggregated origins to the metro stations (as aggregated destinations). The idea is to group the data by the ID of aggregated origins and destinations, and count the trips between each integrated O-D. *Figure 3-11* is an example of the result after the process.

	Id_d	id_station	value
4934	46446	50415	99
15145	28593	50444	70
16656	27832	50448	70
17631	40935	50451	67
10389	26079	50430	50
20600	12030	50460	49
20872	18027	50460	47
20000	24204	50457	46
4882	44989	50415	45
12697	33092	50437	45
461	13261	50401	43
4338	25570	50411	43
816	17806	50401	42
10417	26799	50430	41
818	17808	50401	38
14271	28535	50441	37
17574	40676	50451	37
1303	27851	50402	36
2186	26300	50404	36

Figure 3-11: Aggregated O-D table with the number of trips between each OD ("from metro station trips")

All the steps before are the preparation process for visualization and hotspots extraction. On the one hand, input the tables generated above on the *Flourish* online visualization platform, the flow maps can be created to show the distribution of all bike-metro trips that start from (or stop at) metro stations. On the other hand, after adding coordinates of the new O-Ds by "join" function in *ArcMap 10.7*, the output can be used as the input for the cycling trajectory generation in the next step. *Figure 3-12* is the screenshot of the attribute table in the final step of the process.

start_id	start_x	start_y	stop_id	stop_x	stop_y	value
50415	108.9406	34.36344	46446	108.9435	34.36801	99
50444	108.848	34.28585	28593	108.8425	34.28759	70
50448	109.0647	34.28093	27832	109.0587	34.28324	70
50451	108.9426	34.34243	40935	108.9533	34.34301	67
50430	108.9793	34.27112	26079	108.9794	34.27563	50
50460	108.8435	34.23957	12030	108.8392	34.2126	49
50460	108.8435	34.23957	18027	108.8359	34.23977	47
50457	109.0257	34.27367	24204	109.0283	34.26694	46
50415	108.9406	34.36344	44989	108.925	34.36149	45
50437	109.0278	34.31237	33092	109.037	34.30715	45
50401	108.8687	34.23895	13261	108.8729	34.21803	43
50411	108.9425	34.2716	25570	108.9479	34.27346	43
50401	108.8687	34.23895	17806	108.8566	34.23868	42
50430	108.9793	34.27112	26799	108.9794	34.27889	41
50401	108.8687	34.23895	17808	108.8588	34.23868	38
50441	109.0389	34.27539	28535	109.0402	34.2865	37
50451	108.9426	34.34243	40676	108.9327	34.34193	37
50402	108.8271	34.29124	27851	108.8185	34.28432	36

Figure 3-12: Output table with the location of aggregated ODs

(Note: in this table, the start_x, start_y, stop_x, stop_y are the coordinates of the centroid of the tessellation or buffer.)
All the steps in detail are summarized in the flow chart in *Appendix 5*.

3.2.4.4 Cycling trajectory generation

The cycling trajectory generation process is also in the *Python* environment. As the real routes of the shared bike trips are not available in the dataset, an algorithm based on the Route Plan Application Programming Interface (API) of Baidu Map server will be used to help generate the trajectory based on the origin and destination dataset generated in the previous section. Route plan service (also known as Direction API) is a set of REST-style web service APIs that provides route plan services in the form of HTTP/HTTPS. Currently, the Direction API supports bus, cycling, and driving route planning, and the Direction API supports mainland China (Baidu, 2019).

Import the location of origins and destinations obtained in the previous step according to the rules of the API of Baidu, the optimized trajectory of each O-D will be returned as one csv file that consists of all the trajectory points. In each csv file, several rows represent the trajectory points between the imported O-D pair. The record of each row includes the bike ID, the start and stop time and location of the imported origin and destination. Besides, the duration and linear distance of the imported O-D pair are also included. The "DistanceOnline" and "DurationOnline" shows the distance and riding time of the real cycling routes of each imported O-D pair. The last two columns record the location of each trajectory point in order. It is worth mentioning that the coordinate system of the Route Plan results based on the Baidu coordinate system, and a transformation process is necessary to make the results visualize in the right location. The table below (*Figure 3-13*) gives an example of the scheme of returned csv files. In addition, a *batch point-to-line operation* in the Python environment can group all the trajectory points of each O-D pair into full paths, and the information of each O-D pair is stored in a row of the table. *Figure 3-14* shows the result of the batch processing of "point to line". The flowchart below (*Figure 3-15*) shows all the steps for cycling trajectory generation.

	index	id	start_x	start_y	stop_x	stop_y	STARTTIM	STOPTIME	Duration	Distance	DistanceOnline	DurationOnline	LONGITUDE	LATITUDE
0	1	02900000:	108.992	34.25613	108.9923	34.2535	7:28:00	7:33:00	0:05:00	293.709	479	145	108.99193	34.25611
1	1	02900000:	108.992	34.25613	108.9923	34.2535	7:28:00	7:33:00	0:05:00	293.709	479	145	108.99193	34.25609
2	1	02900000:	108.992	34.25613	108.9923	34.2535	7:28:00	7:33:00	0:05:00	293.709	479	145	108.99193	34.25609
3	1	02900000:	108.992	34.25613	108.9923	34.2535	7:28:00	7:33:00	0:05:00	293.709	479	145	108.992	34.256
4	1	02900000:	108.992	34.25613	108.9923	34.2535	7:28:00	7:33:00	0:05:00	293.709	479	145	108.99203	34.255909
5	1	02900000:	108.992	34.25613	108.9923	34.2535	7:28:00	7:33:00	0:05:00	293.709	479	145	108.99204	34.255889
6	1	02900000:	108.992	34.25613	108.9923	34.2535	7:28:00	7:33:00	0:05:00	293.709	479	145	108.99204	34.255759
7	1	02900000:	108.992	34.25613	108.9923	34.2535	7:28:00	7:33:00	0:05:00	293.709	479	145	108.99204	34.255739
8	1	02900000:	108.992	34.25613	108.9923	34.2535	7:28:00	7:33:00	0:05:00	293.709	479	145	108.99204	34.254879
9	1	02900000:	108.992	34.25613	108.9923	34.2535	7:28:00	7:33:00	0:05:00	293.709	479	145	108.99204	34.254789
10	1	02900000:	108.992	34.25613	108.9923	34.2535	7:28:00	7:33:00	0:05:00	293.709	479	145	108.99205	34.253641
11	1	02900000:	108.992	34.25613	108.9923	34.2535	7:28:00	7:33:00	0:05:00	293.709	479	145	108.99205	34.253501
12	1	02900000:	108.992	34.25613	108.9923	34.2535	7:28:00	7:33:00	0:05:00	293.709	479	145	108.99205	34.253091
13	1	02900000:	108.992	34.25613	108.9923	34.2535	7:28:00	7:33:00	0:05:00	293.709	479	145	108.99205	34.252851
14	1	02900000:	108.992	34.25613	108.9923	34.2535	7:28:00	7:33:00	0:05:00	293.709	479	145	108.99205	34.252651
15	1	02900000:	108.992	34.25613	108.9923	34.2535	7:28:00	7:33:00	0:05:00	293.709	479	145	108.9921	34.252721

Figure 3-13: The result table of "Route Plan"

index	start_id	start_x	start_y	stop_id	stop_x	stop_y	value	DistanceOnline	DurationOnline	LONGITUDE	LATITUDE	geometry
0.0	20588.0	109.010906	34.250634	50400.0	109.008295	34.270913	1.0	3119.0	945.0	109.011158	34.250743	LINESTRING (109.01116 34.25074, 109.01117 34.2...
1.0	21084.0	109.028295	34.252807	50400.0	109.008295	34.270913	1.0	4497.0	1362.0	109.028279	34.252654	LINESTRING (109.02828 34.25265, 109.02760 34.2...
2.0	22027.0	109.009819	34.257154	50400.0	109.008295	34.270913	1.0	2491.0	754.0	109.009794	34.257535	LINESTRING (109.00979 34.25754, 109.01117 34.2...
3.0	22741.0	109.003298	34.260415	50400.0	109.008295	34.270913	1.0	1915.0	640.0	109.003271	34.260384	LINESTRING (109.00327 34.26038, 109.00316 34.2...
4.0	22750.0	109.013079	34.260415	50400.0	109.008295	34.270913	1.0	2225.0	674.0	109.013059	34.260384	LINESTRING (109.01306 34.26038, 109.01270 34.2...
...	...	...	...	...	...	...	...	...	...	...	...	...
3057.0	32507.0	108.922874	34.304974	50461.0	108.942593	34.293570	1.0	3259.0	1017.0	108.922856	34.305090	LINESTRING (108.92286 34.30509, 108.92347 34.3...

Figure 3-14: The results of "points to line"

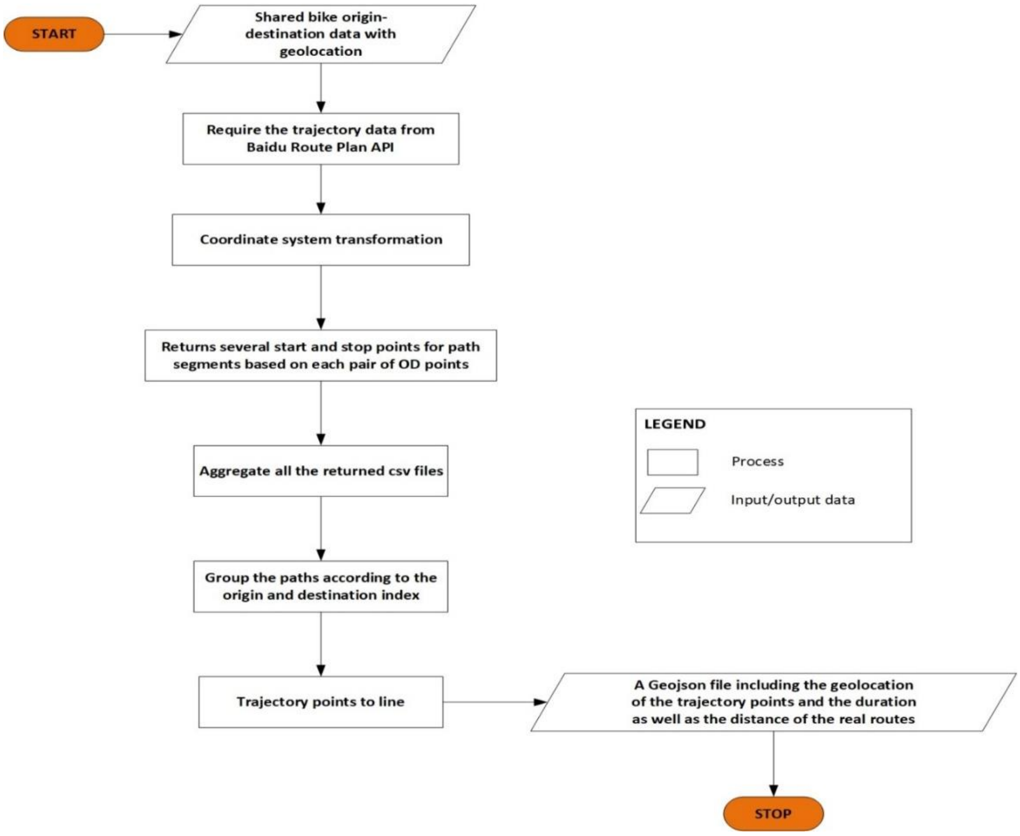


Figure 3-15: Work flow of trajectory generation

3.2.5 Identify the spatiotemporal patterns of shared bike use for metro

This section will analyse the spatiotemporal distribution characteristics of shared bike use. The **first** part is to give an insight to **all the shared bike** use without distinguishing the purposes from both spatial and temporal perspectives. The **second** part will focus only on the **shared bike trips that related to metro**. The proportion of the bike-metro trips, the characteristics of these trips, and the spatial and temporal distribution of these trips will be discussed based on the results of data process.

3.2.5.1 Spatiotemporal characteristics of all shared bike trips

In this section, the analysis focuses on the general spatiotemporal variations of the shared bike trips. First of all, from temporal perspective, all the shared bike trips extracted in the data process are counted in the order of 24 hours a day in Python environment. Then, the temporal variation of the shared bike use in the selected three days can be plotted through the "matplotlib" package. In this way, the peak hours of shared bike use will be determined.

From spatial perspective, the distribution of origin and destination will be plotted through *ArcGIS Pro 2.4.3* to reveal the characteristics of the hot spots of shared bike use in Xi'an city. In ArcGIS Pro, the input data of all the origins and destinations in morning and evening peak hours are displayed through the "*symbolize your layer by density*" function to create heat maps of the imported point data.

3.2.5.2 Spatiotemporal characteristics of bike-metro trips

This section is mainly to analyse the changes in the use of shared bikes connected to the metro. First, to understand the shared bike use as access and egress for metro in a general perspective, the proportion of the bike-metro trips is plotted by *Excel 2016* based on the peak hours and daily dimensions. Besides, according to the introduction of the "Route Plan" API, it is possible to get the real cycle duration and distance. Thus, the histograms based on the result table of "Route Plan" are generated in *ArcGIS pro 2.4.3* to show the distribution of the travel duration and distance of bike-metro trips.

In terms of temporal characteristics of bike-metro use, the visualization software, such as *ArcGIS Pro, Excel* and Matplotlib package based on *Python 3.4* environment will be used to sort data by hours and visualize the temporal variation of usage based on the results of data processing.

On the other hand, from the spatial aspect, **heat maps** through ArcGIS Pro will be used to visually show the hotspots of the distribution of shared bikes use for metro from both daily and peak hour aspects. **Flow maps** that show the direction distribution of the bike-metro trips are also generated based on the result of integrating origins and destinations of bike-metro trips in 3.2.4.3. In this way, the variation of the O-D hotspots of bike-metro trips in different time slots can be revealed so as to give reference to the planners to consider the *parking disorder* and the *rebalancing* challenges as discussed in the previous chapters.

In addition, in combination with the Baidu map "*Route Plan API*" mentioned above, the riding routes of the shared bikes-metro trips will be estimated to a certain extent and visualize on the *Kepler.gl* online platform. As the distribution of the origins and destinations of the shared bikes is very scattered, it is difficult to see the trend of route distribution, using the integrated O-D as the input of visualization platform, the thickness of the line segment can be used to indicate the number of bike trips between the O-D, which is conducive to extract hot spots on the cycling route. In this way, the planners can consider the construction and improvement of cycling facilities, such as *bike lanes* according to the intensity of the cycling routes usage.

3.2.6 Land use - related factors associated with shared bike use patterns

This section will focus on two parts. The **first** part will determine the POI data that can be used to describe the land use. These POIs are selected as indicators that may affect the distribution of the bike-sharing use to connect with metro based on the related literatures and the condition of study area, as well as the data availability. In the **second** part, the indicators selected in the first part will be used as independent variables, and the number of origins (start points) or destinations (stop points) will be used as the dependent variables. These variables as input to do both the global (linear regression) and local regression (GWR). Their association can explain the hot spots distribution of cycling, and also provide a reference for the management of the DLBS system.

3.2.6.1 Determination of independent variables and dependent variables

- **The determination of the dependent variables**

In this research, the number of the origins or the destinations of the bike trips that connect to metro stations is selected as the dependent variables. To explore the relationship between the shared bike use for metro and influencing factors at different geographic locations in the study area, the study area is divided into about 3,000 of 500 meters*500 meters grids. Considering the size of Xi'an City, 500-meter square is used instead of 1km in some relevant studies to provide a more accurate resolution (Yang, 2019; Xu et al., 2019). However, more accurate division will increase the computation time and make it harder to investigate the distribution trend intuitively. The "spatial join" function in ArcGIS is used to add the count of origins and destinations in each cell.

Specifically, for each time slot, both the origins of the "toward metro station" trips and the destinations of "from metro station" trips will be discussed to find out the characteristics between the use of bike-metro trips and the distribution of different land use types.

➢ **The determination of the independent variables**

According to the method of extracting points of interest mentioned in Chapter 3, the table in the *Appendix 2* shows the official classification of all points of interest from the most famous map server, Baidu in China. Comparing the built environment factors used in the relevant literature mentioned in Section 2.4, and the use of POI data as land use discussed in Section 3.2.2, a reclassification is done to select the POIs can be used as built environment factors related to the land use (see *Table 3-4* below). According to these POI types, the link between the shared bike use and the distribution of the different classes of land use can be revealed. The number of each type of POI in each grid will be added to the attribute table through the "spatial join" function in ArcGIS. The POI counts of different types in each cell are the independent variables in the research. The map in *Appendix 6* shows the count of all the selected POIs in each cell.

Table 3-4: Description of the POI reclassification

Category	Type of POI
Job	Companies, factories and government offices
residence	Residential area, dormitory
commercial	Shopping malls, department stores, supermarkets, convenience stores, building materials shop, digital shops, markets, laundry, photo studio, life service shops, beauty salons, barber shop
Recreation	Resorts, cinemas, KTV, theatres, dance halls, internet cafes, game venues, bath and massage, leisure squares, bars
Green	Parks, botanical gardens
Education	Institutions of higher learning, middle schools, elementary schools, kindergartens, adult education, parent-child education, special education schools, scientific research institutions
Health care	General hospitals, specialty hospitals, clinics, pharmacies, medical examination institutions, emergency centers

1. Spatial autocorrelation

In general, spatial autocorrelation is a **prerequisite** for the use of GWR model. If the spatial autocorrelation analysis results of the selected variables are not significantly correlated, the influence of different geographical location on the variables is very weak, and other linear regression models can be used for research instead (Cho, Lambert, & Chen, 2010).The spatial autocorrelation analysis will show positive correlation or negative correlation. When it shows positive correlation, the change trend of unit data value is consistent with that of unit data value of adjacent location; when it shows negative correlation, the trend is opposite (Haining, 2009).

Moran's I (Moran 1950) is often used to identify the overall pattern of **spatial association**. It is used as an overall measure of spatial dependence on the entire dataset (Ester, Kriegel, & Sander, 1999). The characteristic of spatial autocorrelation is the correlation between signals among locations near space. It is complicated because the spatial correlation is multi-dimensional (that is, 2 or 3 dimensions of space) and multi-directional. The calculation formula of Global Moran's I index is as follows:

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} (x_j - \bar{x}) \text{ (Equation 3-1)}$$

Among them, N is the number of features, x is a variable, $\bar{x}$ is the average value of x, w_{ij} is the spatial weight of features i and j; W is the sum of all spatial weights.

The value of I depends largely on the assumption of spatial weights. The idea is to construct a matrix that accurately reflects the assumptions about the specific spatial phenomenon. A common method is that if two regions are adjacent, the weight is 1, otherwise it is 0. Another common method is to assign weights from 1 to k to the nearest neighbours, otherwise 0. An alternative method is to use the distance decay function to assign weights. Sometimes, the length of the common side of the neighbourhood is used to assign different weights to the adjacent areas. The selection of the spatial weight matrix should be guided by the theory of related problems.

The Moran Index is usually between -1 and 1. Moran's I> 0 indicates a positive spatial correlation. The larger the value, the more obvious the spatial correlation. Moran's I <0 indicates the negative spatial correlation. The smaller the value, the greater the spatial difference. Otherwise, Moran's I = 0, means the randomness (CSDN, n.d.).

2. Multicollinearity

In order to reduce the deviation and improve the representativeness of the coefficients in the model, the multicollinearity between the potential independent variables is tested. The variance inflation factor (VIF) is used to test the multicollinearity. The higher the VIF value, the more serious the effect of multicollinearity. Since there is no fixed threshold, it is generally considered that there is a high multicollinearity between independent variables when the VIF > 5. On the basis of this experience, this study uses *IBM SPSS Statistics 25* software to do regression analysis for the available independent variables, and obtains the collinear diagnosis results of each variable.

3.2.6.2 Application of geographic weighted regression model in this study

The geographic weighted regression model combined the geographical position of the data into the regression parameters. GWR Model is processed according to the following rules: the closer the geographical location is, the greater the impact on the results.

Brunsdon, Fotheringham and Charlton (1996) introduced the term GWR, which refers to a series of "spatially adjusted" regressions that work by assigning weights to each observation. The regression depends on the distance from a specific geographic location or focal point (Páez and Wheeler, 2009). The analysis involves fitting as many regressions as the observations, and it is based on the concept of distance decay. These weights are generated by a kernel function. This kernel function uses the bandwidth found by optimizing the goodness-of-fit criterion. As a result, the estimated value can be fitted to each observation or position by applying an appropriate equation.

1. The weight function in the GWR model

The GWR, like other regression model, should first define a research area. Then the most important thing is to use the different spatial location of each feature to calculate the decay function to generate the corresponding weight according to the geographical location of the feature. The decay function can be used to calculate all sample points. In the calculation of each sample point, different weights are given to samples according to the different spatial relationship with the sample point, so as to obtain the regression coefficient of each sample point. Finally, all the regression coefficients can be interpreted.

The weight functions of GWR Model are mainly divided into two categories: one is **fixed** weight function, that is, the weight function is fixed in the whole research area. The other more widely used one is **adaptive** weight function, which can be adjusted according to the density around the calibration points. In spatial analysis, Kernels are often used as weighting functions to model and explain spatial autocorrelation and heterogeneity (Wang & Cheng, 2012). The **bi-square kernel** function is the most commonly used one as it is effective to eliminate the sample points that are far away and have almost no effect to improve the efficiency of parameter estimation (Bidanset & Lombard, 2014). Thus, in this study, a default kernel type of **adaptive bi-square** will be used in the analysis.

2. Bandwidth in the GWR model

Previous studies have shown that compared with the selection of spatial weight function, it is more significant to determine the **bandwidth** of spatial weight function. A larger bandwidth will increase the deviation of parameter estimation, and the difference of parameter estimation in different regions will be small. A smaller bandwidth will reduce the effectiveness of parameter estimation and increase the variance. Therefore, the method to determine the optimal bandwidth has been widely studied. At present, there are three commonly used methods: cross validation (CV), Akaike information criterion (AIC) and Bayesian information criterion (BIC)(Khoi & Murayama, 2012). *AICc* is usually used as a correction for small sample sizes. Since AICc takes into account both simplicity and accuracy when evaluating the model, the optimal bandwidth is determined according to AICc in this research.

3. The analysis process of GWR model

GWR can be viewed as an extension of the general linear regression model. The GWR model allows coefficients to vary with space (Lu, Charlton, Harris, &Fotheringham, 2014). Compared with the general linear regression model, GWR usually has a smaller residual and a smaller spatial dependence. Therefore, the GWR model generally based on the global regression model. From other relevant research, the linear regression based on Ordinary Least Square (OLS) is the best known and most common one of all regression techniques. It is also the proper start for the GWR (Esri, n.d.-a). OLS model is usually used to determine the global regression coefficient of independent variables (β). Once the independent variables are identified in the model, there is a theoretical basis to think that the relationships may vary from geographic location, and the GWR is potentially the appropriate next step (Columbia, n.d.).

The GWR model explicitly considers the spatial components of the data and incorporates the geographic coordinates of the observations into its equations. Unlike the OLS model, in the GWR model, the coefficient βj ($j = 0,1, ..., p$) of the variables χj varies with the location. For each location defined by its coordinates (u_i, v_i), the value of the dependent variable y_i is estimated according to equation (3-2).

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_1 + \beta_2(u_i, v_i)x_2 + \cdots + \beta_p(u_i, v_i)x_p + \varepsilon_i \quad \text{(Equation 3-2)}$$

In this research, after determining the dependent variable and the potential independent variables, first, the linear regression based on the Ordinary Least Square function in *ArcMap 10.7* software will be used to conduct a preliminary analysis of the relationship between the independent variables and the dependent variable, so as to identify the suitable independent variables that can be used to study spatial heterogeneity. Then, after considering the necessity of using the GWR model according to the distribution of the residual squares in the OLS model, input the independent variables into the GWR model in the *GWR 4.0* for final analysis. Then, the process is to assess the benefits of moving from the OLS model to a local regression model (GWR) by comparing the R^2 and the value of Residual Square, as well as the AICc between these two models (Esri, n.d.-a). Through such a research process, not only can the factors affecting the distribution of bike-metro use be determined, but also the different impacts of these factors on the independent variables caused by different geographic locations within the study area will be revealed. The flow chart below summarizes this research process (*Figure 3-16)*.

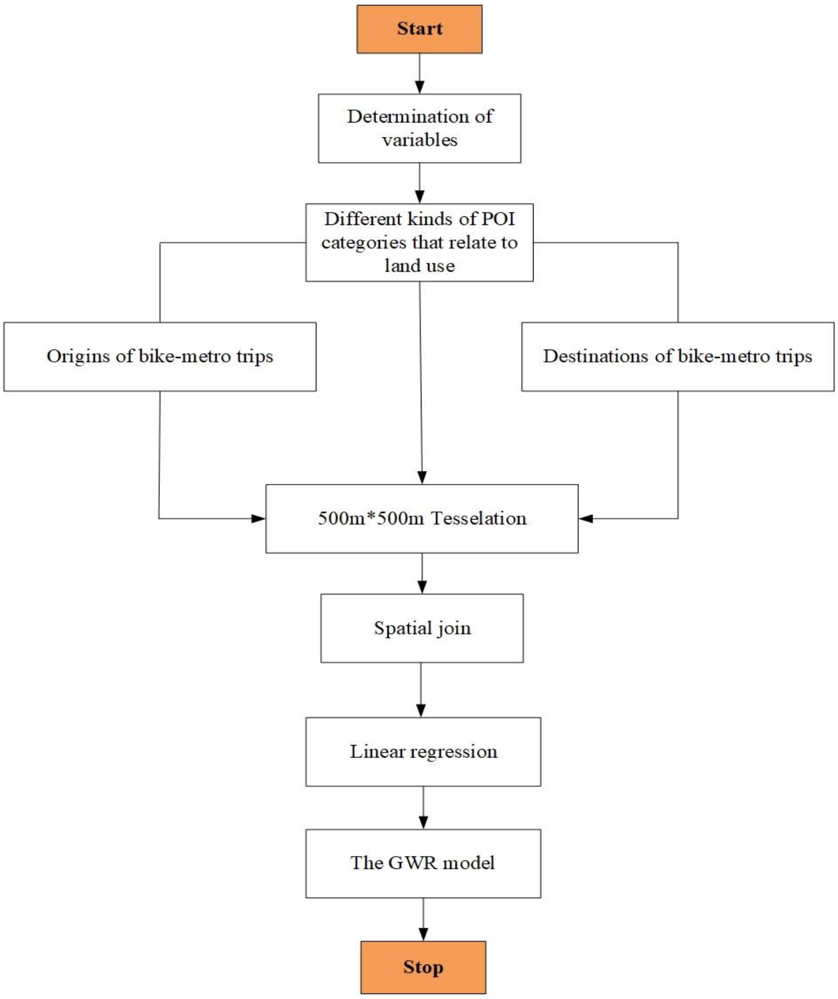


Figure 3-16: Work flow for the GWR model

4 RESULTS AND DISCUSSION

In this section, all the visualization outputs based on the data processing are displayed and discussed to help understand the spatiotemporal characteristics of the shared bike use for metro system, and the factors that affect the use patterns.

Three main subsections are included in this chapter. First, the overall use patterns of shared bike in Xi'an city are discussed without identifying whether the bikes are used for connecting to the metro system. In this way, a general idea about the bike-sharing usage can be provided, which is also helpful to prove the relevance of analyzing the bike-metro trips in study area. Second, according to the method presented in section 3.2.4.2, all the bike-metro trips can be identified in the study area. Thus, the spatiotemporal patterns of the shared bike use for metro can be visualized from several perspectives. The analysis can effectively help to observe the hotspots of bike-metro trips from both a temporal and a spatial perspective. Third, a set of urban land use characteristics is analysed to identify if there is a relation with the intensity of shared bike usage during peak hours.

Combining the temporal distribution and the spatial patterns of shared bike use in three workdays *(Appendix 7)*, which are also quite similar, thus, only the use patterns on the 3rd of September (Monday) are analysed in detail to develop an approach to investigate the spatiotemporal patterns of shared bike use. Moreover, according to the study of the daily variation in the use of shared bikes, it is found that the use patterns is the most intensive in the morning and evening peak hours, and there are very different characteristics. Therefore, the focus will be on the distribution of shared bike use from the morning and evening peaks on September 3.

4.1 Spatiotemporal characteristics of all shared bike trips

4.1.1 Temporal patterns of the shared bike use

First, based on the Python 3.4 environment, the usage of all the shared bikes over time in the selected three days was counted and displayed with a line chart (*Figure 4-1*). The usage patterns during three days remained basically stable. Moreover, it can be clearly distinguished that the two peak periods of bike use are 7: 00-9: 00 in the morning and 17: 00-19: 00 in the evening, which is consistent with the commuting time of many commuters in Xi'an.

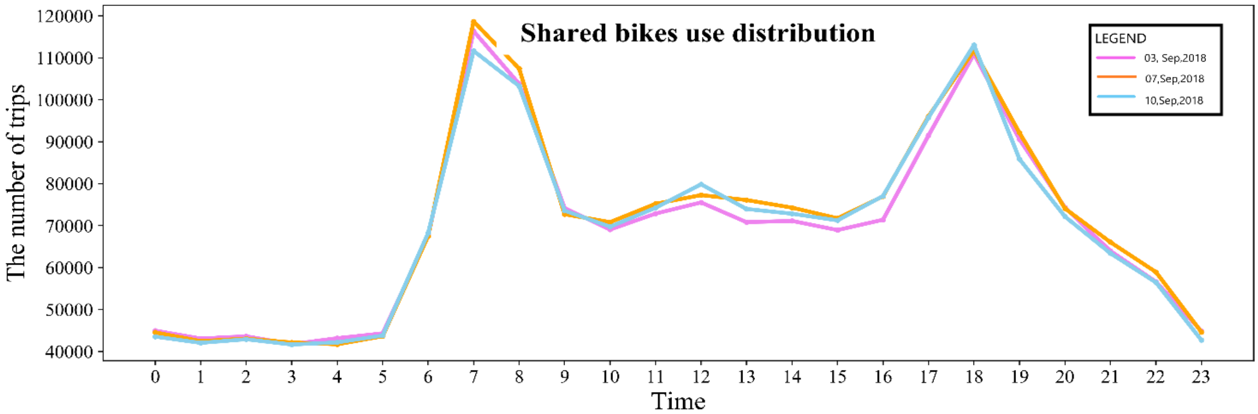


Figure 4-1: The daily temporal distribution of shared bike use

4.1.2 Spatial patterns of the shared bike use

In this section, the overall distribution of all the origins and destinations of shared bike use were plotted based on the extracted O-Ds. The heat maps in *Appendix 7* show the distribution of the origins of the bike-sharing trips in three days. However, the spatial pattern of bike-sharing use in three days is very similar. The temporal usage of shared bikes during three days is also very similar from the previous analysis. Therefore, in the following research, the focus of the analysis will be **only** on Monday, **September 3rd**.

It can be seen from the maps in *Figure 4-2* that the areas with the densest distribution of origins and destinations are all concentrated in the southwest part of the city. This can be associated with the administrative division of Xi'an. The southwest region is the High-tech Development Zone in Xi'an. There are many employment communities, and residential areas, and the density of metro lines in this area is low. Thus, there is more demand for bike-sharing.

On the one hand, in the **morning** peak hours, the usage distribution of the origins of the shared bikes is more scattered than the distribution of destinations, and it is widespread along the metro lines. Because the hot spots of the origins of cycling are mostly distributed in the dense residential areas of the high-tech zone, and the destinations of cycling are mostly job-related locations. In addition, in the high-tech zone, the distribution of residential areas is more scattered than the jobs. On the other hand, in the **evening** rush hour, the hotspots of the destinations of shared bikes spread around the metro lines, and the diffusion coverage is obviously wider than the patterns of origins of the trips, indicating that people use shared bikes to have a more diverse of activities at evening peak.

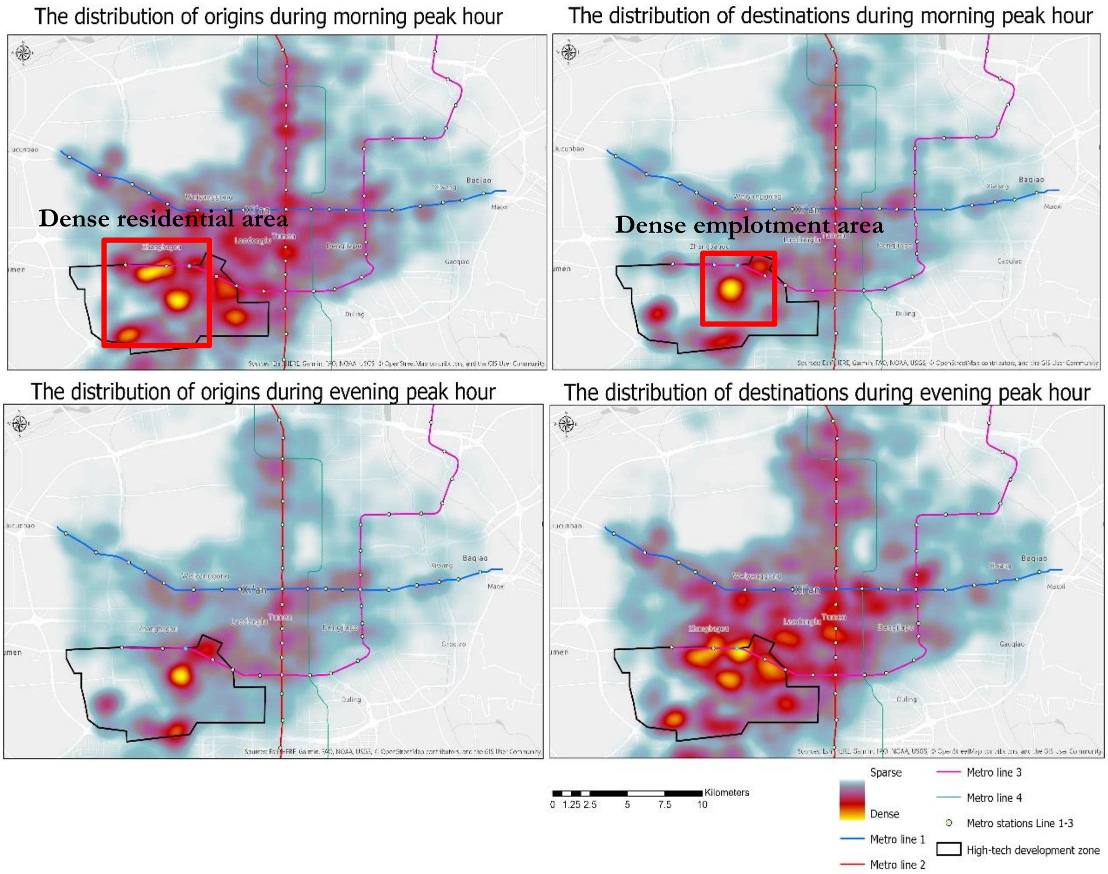


Figure 4-2: The distribution of the start and stop points of all shared bike trips in peak hours

4.2 Spatiotemporal characteristics of bike-metro trips

In this section, all the visualizations focus **only** on the riding trips that connect with the metro system on 3rd of September. There are three subsections included: **first**, the overall usage of bike-metro trips is revealed from both the proportion and the characteristics to have a general idea about the usage of bike-metro trips. **Second**, the daily variation of the bike-metro trips is described through graphs to identify the peak hours and off-peak hours of the shared bike use for metro system. **Third**, the spatial distribution patterns of the bike-metro use are visualized in different ways to find out the hotspots of the shared bike use for metro system.

4.2.1 Overall usage of the bike-metro trips

4.2.1.1 The proportion of bike-metro trips

The pie chart below gives an idea about the average proportion of the shared bike trips to connect with metro system in three workdays. About 15% of the bike-sharing trips are metro-related. The proportion becomes even higher in the morning peak hours, which is about one fifth. In contrast, the share of metro-related bike trips during the evening rush hour dropped to the average. However, the ratios are also sufficient to show that quite a few of travellers use shared bikes as access or egress to metro stations.

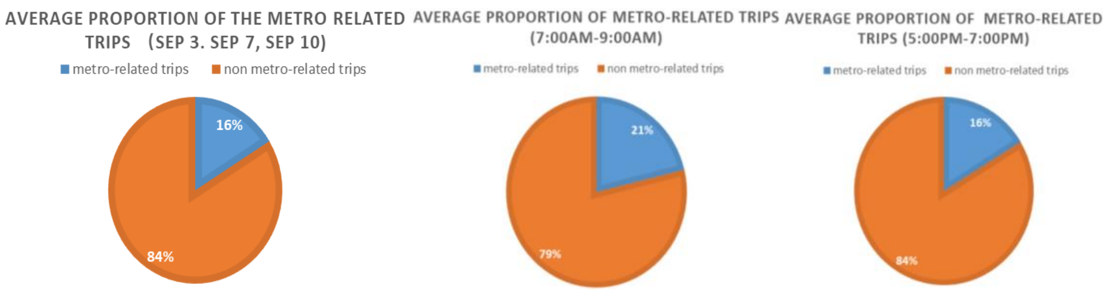


Figure 4-3: The proportion of shared bike use for metro

4.2.1.2 The distribution of travel distance and duration of bike-metro trips

The histograms in *Figures 4-4* and *Figure 4-5,* generated in **ArcGIS pro 2.4.3**, show the distribution of the cycling distance and the travel duration between the origins and destinations of the bike-metro trips on September 3. Both the distance and the travel duration based on the algorithm of the "route plan" API from Baidu Map server. The returned values include the distance and time of the cycling routes based on the real road network.

The mean of distance is 1674 meters, and the mean travel duration between the origins and destinations is about 9 minutes. The use frequency of shared bikes to connect with metro shows a tendency to rise within a distance around 1200 meters or travel time around 6 minutes, but declines over a certain travel distance or time. In general, 90% of the users spent less than 15 minutes or a short distance around 3000 meters to finish the bike trips that connect to the metro stations.

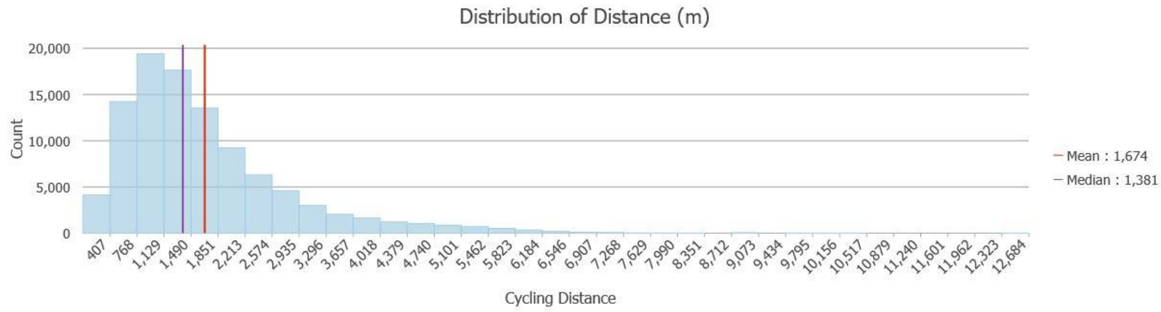


Figure 4-4: The distribution of linear distance of bike-metro trips

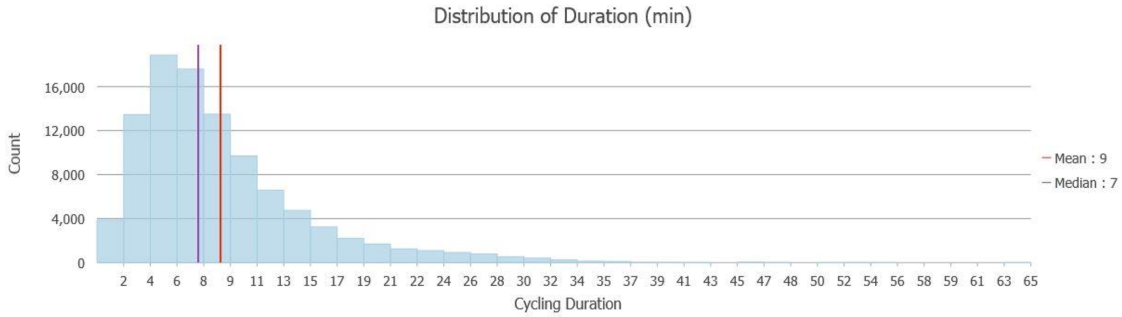


Figure 4-5: The distribution of travel duration of bike-metro trips

4.2.2 Temporal patterns of bike-metro trips

The graphs below show the temporal variation of bike-metro trips during a day. The same as the temporal variation of all shared bike usage discussed before, the temporal patterns of bike-metro trips are quite similar among three days. Unlike the use patterns of all shared bikes, the use frequency of bikes that connect to metro stations in morning peak is obviously higher than that in the evening peak. There is a sharp decrease after 7:00, which is also different from the trend of all shared bike trips. The number of trips from 17:00-18:00 is increasing, but a steep drop occurs after 19:00 in the evening peak.

It can be seen from the temporal patterns and the overall distribution characteristics of bike-metro trips in the previous section that the three-day bike-metro usage is very similar. In addition, there is a big difference in the usage of shared bikes between morning / evening peak and off-peak hours. Besides, the bike-metro usage intensity is very similar to the overall intensity of all the shared bike usage, which means there are lots of people use the shared bikes for connecting with metro system, as well as for other purposes in peak hours. In order to consider the usage distribution specifically, the following analysis will pay more attention on the use of shared bikes **in the morning and evening peak hours**.

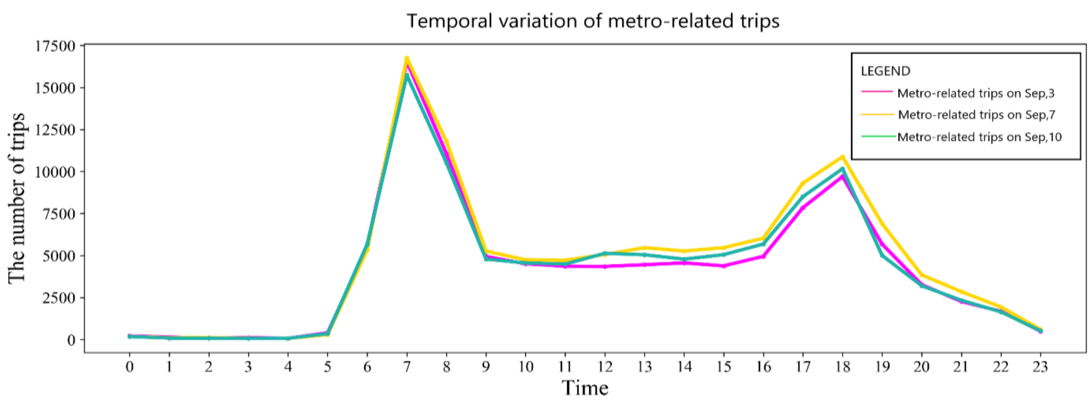


Figure 4-6: Temporal variation of the shared bike trips to connect metro system

4.2.3 Spatial patterns of bike-metro trips

4.2.3.1 The distribution patterns of the origins and destinations of bike-metro trips

The distribution patterns of the O-Ds of bike-metro trips is analysed mainly from two aspects based on the trip directions. One is the shared bike trips start from metro stations and ride to other destinations, the other one is the shared bike trips start from other origins and ride to metro stations. *Figure 4-7, Figure 4-8* below combined the **morning and evening peak hours** from the temporal dimension to consider the distribution of the origins and destinations of the bike trips.

First of all, there are some clear spatial clusters of the O-D in peak hours. From the perspective of the distribution patterns of "**toward metro station**" trips **in morning rush hours**, there is a hot spot near Yuhuazhai Station on Metro Line 3 as it is a residential area. The area near transfer station Beidajie also generated more cycling trips, as the development intensity of this area is advanced with high degree of land use mix. There are also many job-related and residential areas. Thus, these areas are hot spots for both the origins of "to metro station" trips and the destinations of "from metro station" trips. Similarly, there are also some hotspots locate near Tonghuamen station. From the perspective of "**from metro station**" trips, some clusters of the destinations distribute from the north of Xingzhengzhongxin station to the south of Xiaozhai Station on the Line 2. Among them, Xingzhengzhongxin station is the administrative center in the city, and there are different levels of governmental offices. Thus, it is also a hot spot that attracts many commuters cycling from metro stations. On Line 3, Kejilu, Taibainanlu, and Yanpingmen Station, are also parking hot spots of shared bikes, because of the concentration of employment places in the high-tech zone.

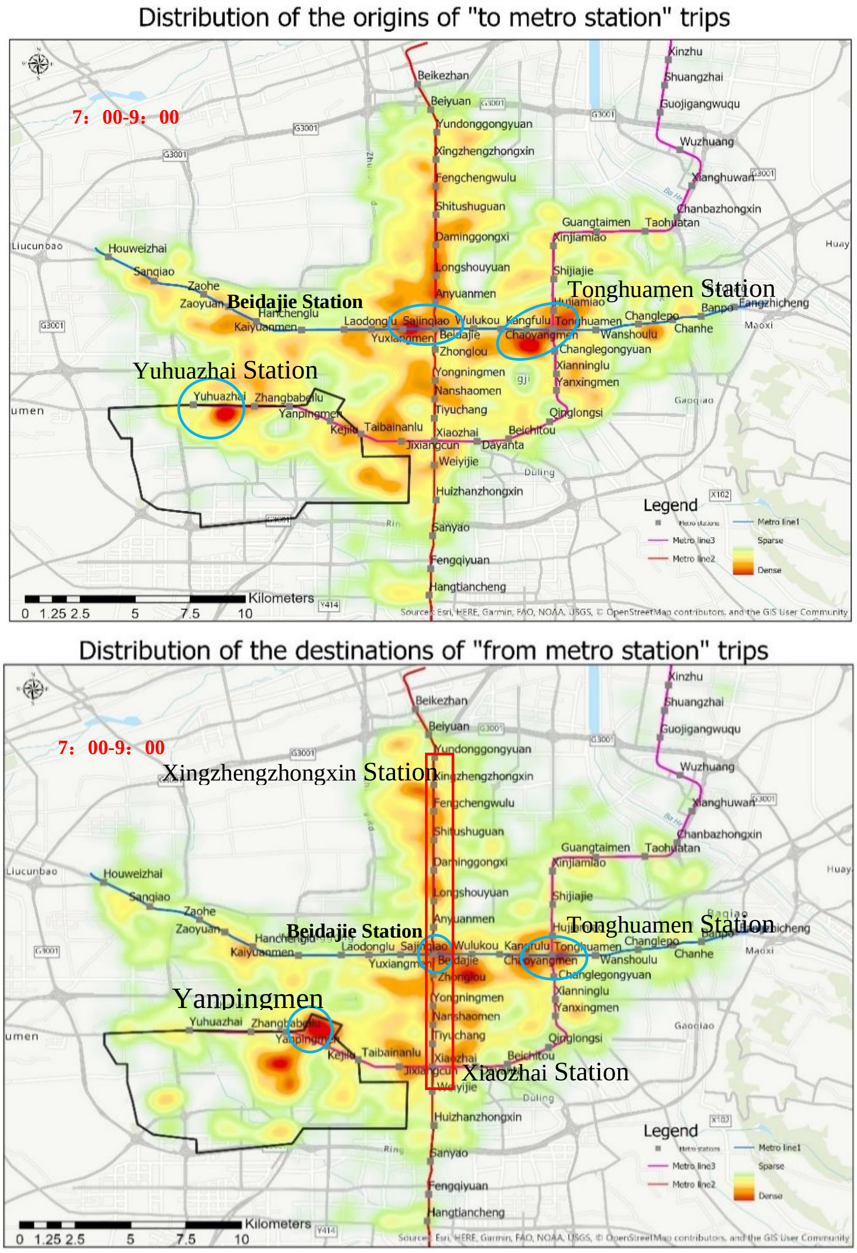
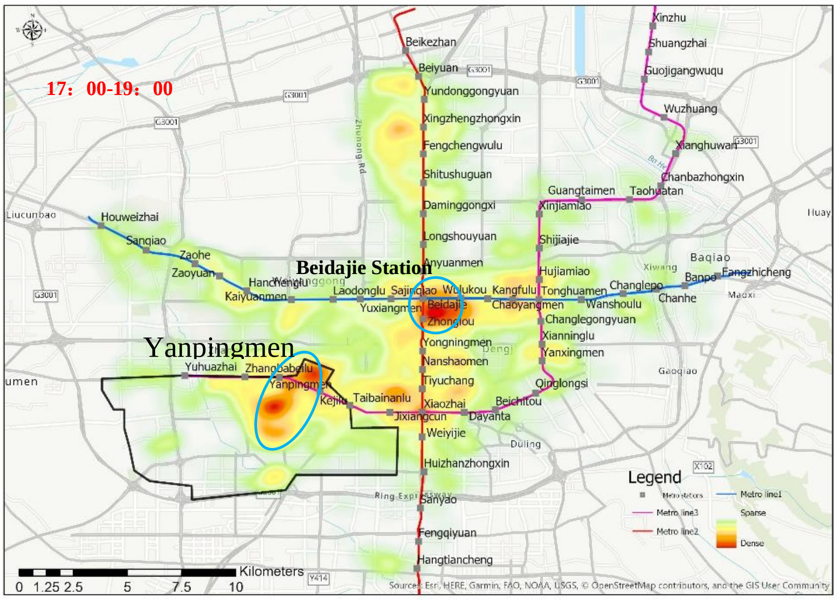


Figure 4-7: Distribution of the origins and destinations of "to metro station" and "from metro station" trips in the morning peak hours

During the **evening peak hours**, the hotspots of the **origins** decreased significantly compared with morning peak as the hot spots mainly distribute in the high-tech area, which is one of the densest employment areas in the city. Only two obvious clusters can be seen on Beidajie station, and the south of Yanpingmen station. Moreover, the distribution of **destinations** of the bike-metro trips that start from the metro stations has a tendency to expand outwards more compared to the morning peak, which indicates that the users depart from the metro stations has a longer riding distance, and the destination of **the riding activity is more varied in the evening peak**. In addition to the hotspot areas that appear in the morning rush hours, hotspots also appear in the south of Changlepo Station on Line 1, Hujiamiao Station on Line 3, as well as the areas in the south of line 3 from Yuhuazhai station to Zhangbadonglu station. These are all the residential concentrated areas, which indicates users are more likely to ride shared bikes to go home from metro system after work.

Distribution of the origins of "to metro station" trips



Distribution of the destinations of "from metro station" trips

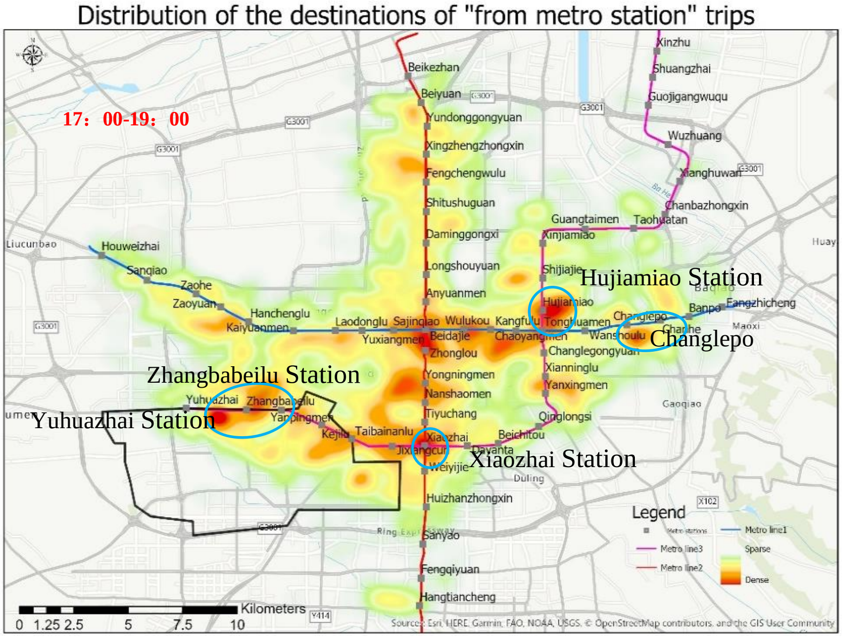


Figure 4-8: Distribution of the origins and destinations of "to metro station" and "from metro station" trips in the evening peak hours

On the other hand, the interactive flow maps were generated on "Flourish" visualization platform to show the similar information of the O-D distribution. However, the heat maps of O-Ds cannot show the riding directions of from or to metro stations. Therefore, the flow maps in *Figure 4-9, 4-10* are used to indicate in which directions the riding trips from or to the metro stations. The red end of each flow line represents the origin, and the blue end represents the destination. The thickness of the flow line represents the number of trips between the O-Ds, and the thicker the flow line, the more trips occur between the O-D pair. The riding trip from the metro station can be regarded as the trip generated by the station, otherwise it is the trip attracted by the station. Therefore, it is intuitive to see the generation and attraction patterns of bike trips in each metro station through the flow map. Combining the O-D heat maps above, it is useful to identify the variation of the hot spots of O-Ds during peak hours, which is helpful to provide evidence to deal with the challenges such as **parking disorder and rebalancing** that caused by the development of DLBS.

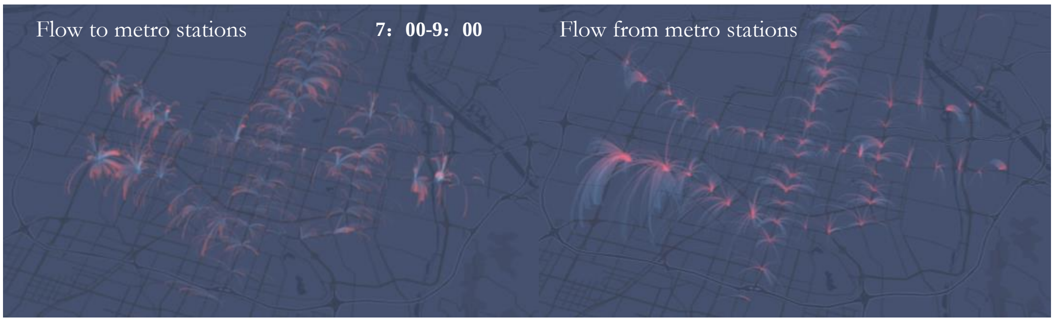


Figure 4-9: The flow map of bike-metro trips during morning peak hours
(https://app.flourish.studio/story/232246/edit)

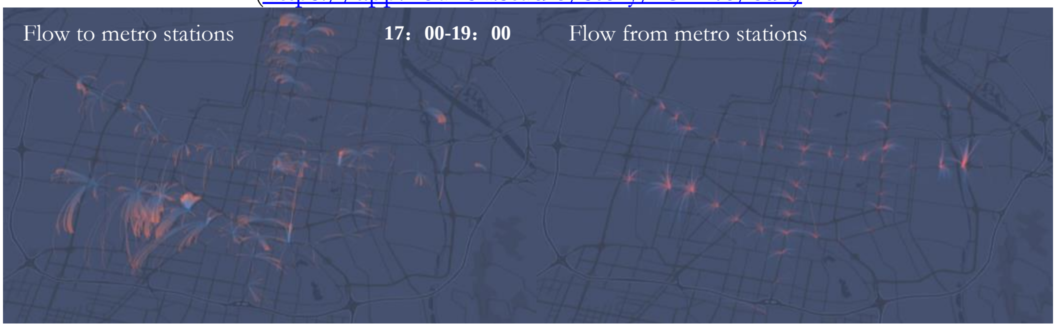


Figure 4-10: The flow map of bike-metro trips during evening peak hours
(https://app.flourish.studio/story/233467/edit)

4.2.3.2 The distribution patterns of the riding routes of bike-metro trips

In order to understand the spatial characteristics of the bike-metro trip more comprehensively, the cycling route intensity can help identify the most popular routes, thereby providing a reference for **bike lane** planning and improvement. Based on the origins and destinations aggregated to the 100-meter grid, the distribution of riding routes is visualized according to the rush hour time slots. The thickness of the riding route will be used to indicate the number of trips on the route.

From *Figure 4-11*, in the morning and evening rush hours, the route on the south of Zhangbabeilu Station on Line 3, and the third ring road on the south of Yuhuazhai Station are very frequently used routes of bike trips from metro station in the morning. They are also popular routes of the trips toward metro stations in the evening. Because the destinations or the origins of these paths correspond to the concentration of jobs, such as software development companies (See *Appendix 6*). In addition, the highly used riding routes of "from metro station" trips in the **morning rush hour** are mainly distributed near the Xingzhengzhongxin Station of Line 2, which is also a concentration of governmental offices. Looking at the **evening peak hour**, the distribution of cycling routes hot spots in both directions shows the trend of distribution opposite to the morning peak. Besides, a new hot spot for cycling routes appeared near Beiyuan Station on the northern of Line 2, which is a residential area with a long-distance bus station. Because the users are likely to ride from job concentrations to the metro stations, or ride from metro stations to residential concentrations after work during evening peak hours. It is worth noting that, compared with the morning rush hour, the usage of the routes in the high-tech zone around Kejilu Station, Taibainanlu Station has increased significantly in the evening peak. This shows that during the evening rush hours, many cycling trips in the high-tech zone are not job-related. The purposes of bike-metro trips in the evening rush hours are more diverse. In short, the distribution patterns of cycling routes in the morning and evening peak hours has a highly consistence with the **commuting activities**.

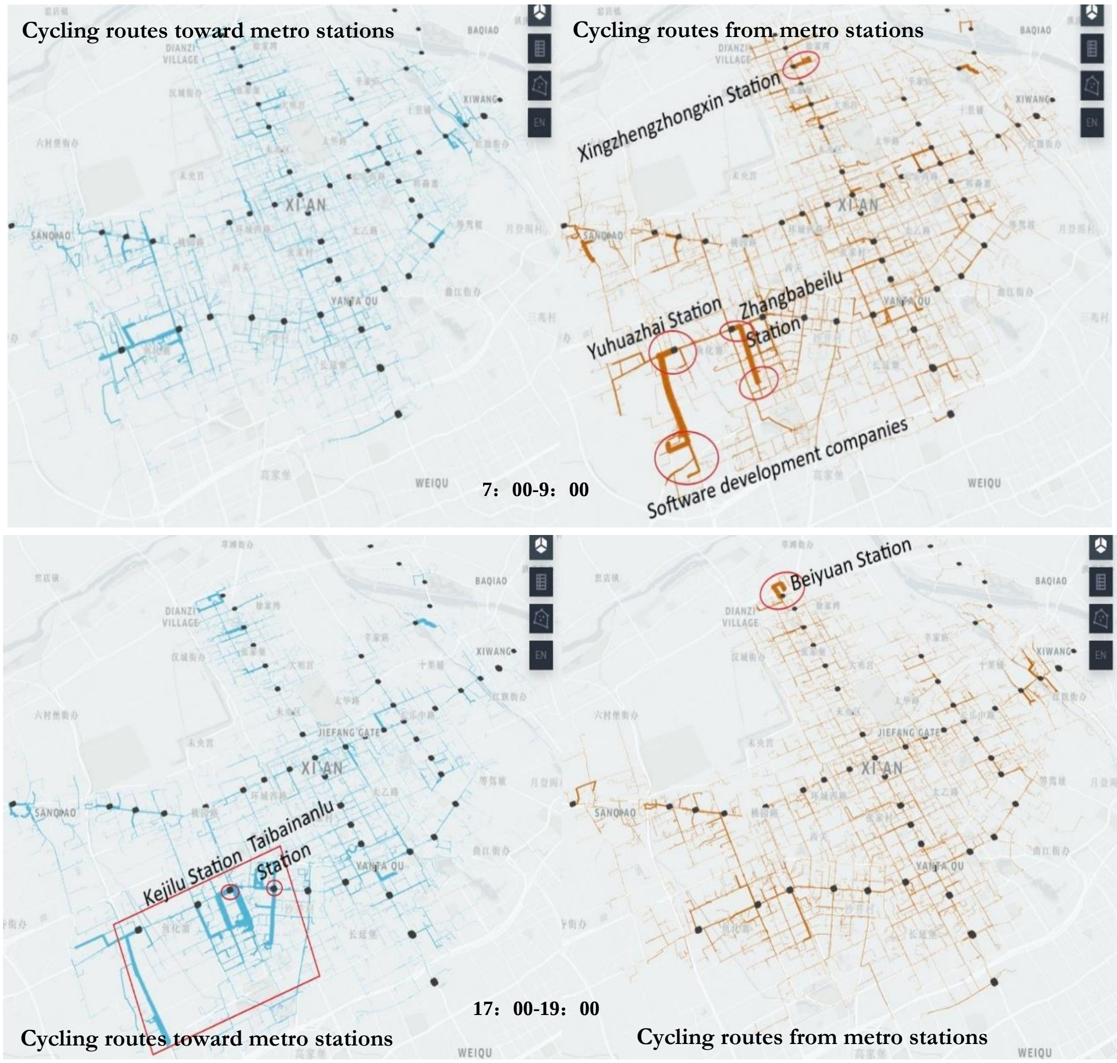


Figure 4-11: The intensity of the riding routes during peak hours

4.3 Influencing factors of shared bikes use for metro

According to the analysis above, the usage of shared bikes in the morning and evening peaks is significantly large, and there are obvious differences in distribution characteristics. Thus, in this section, the link between the POI-based and land use related factors and the shared bike use for metro will be discussed according to different time slots of peak hours, which are 7:00-9:00 in the morning, and 17:00-19:00 in the afternoon. During these hours, all the shops and services are open. Specific analysis is according to the following *Table 4-1*.

Table 4-1: Time slots and contents

Time slots	Dependent variable
7:00-9:00	The number of Origins of "toward metro station" trips
	The number of Destinations of "from metro station" trips
17:00-19:00	The number of Origins of "toward metro station" trips
	The number of Destinations of "from metro station" trips

➢ **Multicollinearity of the variables**

The VIF value of each explanatory variable is less than 2 means there is no multicollinearity between the variables. See global regression results below for detailed diagnosis results.

➢ **Spatial autocorrelation test**

The results of Global Moran's I test for all the potential independent variables show in the table (*Table 4-2*) below. The *expected index* means the expected value of Moran's I under the null hypothesis of no spatial autocorrelation. Thus, all the values of Moran's I are higher than the expected index show the positive spatial autocorrelation exist. The *z-score* is a test of statistical significance that helps to decide whether or not to reject the null hypothesis. The *p-value* is the probability that you have falsely rejected the null hypothesis (Esri, n.d.-c). From the results, the p-value is close to 0, and the z-score values are all positive showing statistically significant, the spatial distribution in the dataset is spatially clustered. This result implies that the difference in geographic location has a greater impact on the variables. Thus, the GWR model have more potential to explain the spatial heterogeneity of these variables (Bao, Liu, Yu, & Xu, 2017).

Table 4-2: Global Moran's I test for potential independent variables

Variables	Moran I	Distribution	z-score	p-value	Expected Index
Job	0.448990	clustered	34.592435	0.000000	-0.000337
Residence	0.551428	clustered	42.212100	0.000000	-0.000337
Commercial	0.503600	clustered	38.529894	0.000000	-0.000337
Recreation	0.509284	clustered	39.011748	0.000000	-0.000337
Green	0.463157	clustered	36.592806	0.000000	-0.000337
Education	0.253419	clustered	19.531937	0.000000	-0.000337
Health care	0.357438	clustered	29.950289	0.000000	-0.000337

4.3.1 Global regression analysis (OLS model)

Before using the GWR model to explore the spatial heterogeneity of the distribution of influencing factors on the origins and destinations of the shared bike trips connected to the metro, a global regression model is used to perform regression analysis on the dependent variables and independent variables to give a **general idea** about how the factors influence the dependent variables, so as to determine the impact factors. Specifically, linear regression based on the ordinary least square (OLS) model will be operated to the POI-based variables and the origins and destinations of bike-metro trips according to different time slots.

4.3.1.1 OLS model results for the morning peak hours

(1) The number of origins as dependent variable

The OLS model was carried out on the origins of bike trips "toward metro station" from 7:00 to 9:00 in the morning, as well as the influencing factors. Looking at the regression results in *Table 4-3*, the R^2 indicates that the independent variables can explain about 31% of the variation of the dependent variable. The p value less than 0.05 (significance level $\alpha = 0.05$) shows the high statistical significance of the model (DSS, 2007). Thus, the number of residences, commercial, recreation, and healthcare POIs have a significant linear correlation with the origins of the bike-metro trips in the morning peak. Other variables have no association with the origins of the bike-metro trips in the morning peak. Therefore, we remove these variables for further regression.

Table 4-3: The results of the OLS model for all independent variables

Variables	Std. Coefficients	p-value	VIF
Intercept		0.000	
healthcare	0.203	0.000	1.388
education	-0.005	0.867	1.306
green	-0.046	0.078	1.121
recreation	0.113	0.001	1.914
commercial	0.201	0.000	1.592
residence	0.205	0.000	1.259
job	0.037	0.214	1.497
R^2	**0.306**		

The standardized beta coefficients compare the strength of each independent variable's influence on the dependent variable. From the new results of regression (*Table 4-4*), all the standardized coefficients of the variables are positive, which means the number of residences, commercial, recreation, and healthcare POIs are positively correlated with the distribution of the origins of the bike sharing trips during the morning peak. Thus, the increase in these land use types will increase the use of shared bikes.

Specifically, the **standardized coefficient** value of variable "residence" is 0.213, indicating that with every increase of one standard deviation in the number of residences, the number of origins of bike-metro trips in the morning rises by 0.213 standard deviations. Therefore, residences have the greatest impact on the distribution of the origins of bike-metro trips in the morning peak hours, and the impact of recreation is small.

Table 4-4: The results of the OLS model for selected independent variables

Variables	Std. Coefficients	p-value	VIF
Intercept		0.000	1.375
healthcare	0.209	0.000	1.304
recreation	0.103	0.001	1.479
commercial	0.203	0.000	1.238
residence	**0.213**	0.000	1.375
R^2	**0.303**		

(2) The number of destinations as dependent variables

Using the same method, the independent variables that affect the distribution of the destinations of bike trips "from metro stations" in the morning peak are the number of job locations, recreation, healthcare, education, and commercial areas. From the table (*Table 4-5*) below, the R^2 increased to 0.480 when the number of destinations of bike-metro trips was used as dependent variables. It indicates that the model has a good fitness that 48% of the variance in the dependent variable can be explained collectively by the independent variables. According to the standardized coefficients, the number of job locations has the greatest impact on the distribution of the destinations of shared bike trips in the morning. In short, we can conclude that the more job locations, the more bike trips, while the increase of the number of commercial areas will slightly reduce the distribution of the destinations of shared bike trips in the morning. Besides, the amount of recreation POIs and education POIs also have an influence on the distribution of the destinations of the bike-sharing in the morning peak, and the degree of influence gradually weakens.

Table 4-5: The results of the OLS model for selected independent variables

Variables	Std. Coefficients	p-value	VIF
Intercept		0.000	
healthcare	0.169	0.000	1.362
education	0.079	0.002	1.311
recreation	0.175	0.000	1.668
commercial	-0.063	0.023	1.461
job	**0.395**	0.000	1.186
R^2	**0.479**		

4.3.1.2 OLS model results for the evening peak hours

(1) The number of origins as dependent variables

The table below (*Table 4-6*) shows the results of performing a second linear regression after removing variables that do not meet the conditions. From the standardized coefficients, the number of "job" POI has a significant positive effect on the distribution of the origins of the shared bike trips during evening peak hours. The **standardized coefficient** value of variable "job" is 0.430, which means with every increase of one standard deviation in the number of jobs, the number of origins of shared bike trips in the evening increases by 0.430 standard deviations. In short, the denser of the job locations, the more bike trips start from the area to the metro stations. Then, the number of recreation POI also effects the distribution of the origins of bike-metro trips, but the impact is not very strong. Besides, the number of healthcare, educational buildings all have a weak positive effect on the dependent variable. The commercial has a slightly negative impact on the dependent variable, which means with every increase of one standard deviation in the number of commercial areas, the number of origins of shared bike trips in the evening decreases by 0.107 standard deviations.

Table 4-6: The results of the OLS model for selected independent variables

Variables	Std. Coefficients	p-value	VIF
Intercept		0.000	
commercial	-0.107	0.001	1.664
healthcare	0.120	0.000	1.315
education	0.085	0.003	1.298
recreation	0.226	0.000	1.669
job	**0.430**	0.000	1.273
R^2	**0.360**		

(2) The number of destinations as dependent variables

From the coefficient value in *Table 4-7*, it can be seen that these four variables all have a positive effect on the distribution of the destinations of the shared bike trips in the evening peak hours. Among them, the number of recreation POIs, residential POIs, healthcare POIs and job POIs have a significant impact on the dependent variable, and the degree of impact decreases in turn.

Table 4-7: The results of the OLS model for selected independent variables

Variables	Std. Coefficients	p-value	VIF
Intercept		0.000	
Healthcare	0.162	0.000	1.242
Recreation	**0.273**	0.000	1.509
Residence	0.182	0.000	1.187
Job	0.141	0.000	1.209
R^2	**0.307**		

Through the above analysis, the significant influencing factors of each time slot are summarized in the following table (*Table 4-8*), the **bold variables** are the most influential independent variables. From the results of global regression, the more residences, the more origins of bike-metro trips, while more job locations will attract more bike trips that start from metro stations in the morning peak hours. Meanwhile, in the evening peak hours, the distribution of the origins of the bike-metro trips are positively influenced by the number of jobs, while the distribution of the destinations of bike-metro trips are positively influenced by the number of recreation areas. The results are consistent with the general perceive of the commuting activities. The commuters ride from home to metro stations, and take metro to work or ride from metro stations to job locations in the morning peak. In the evening peak hours, the shared bike users ride bikes from job locations to metro stations to go home or from metro stations to recreation areas to relax.

Table 4-8: Summary of the global regression results

Time slots	Dependent variables	Independent variables
Morning peak (7:00-9:00)	Origins	**Residence**, commercial, recreation, healthcare
	Destinations	**Job**, commercial, recreation, education, healthcare
Evening peak (17:00-19:00)	Origins	**Job**, commercial, recreation, education, healthcare
	Destinations	Job, residence, **recreation**, healthcare

The map below (*Figure 4-12*) plotted the residuals of the OLS model to enable identification of areas where the global regression fails to explain a large proportion of the variation (Lloyd, 2010). From the maps, the negative value of the residuals represents the model is underestimating, the positive value shows the model is overestimating. Thus, both the blue and red cells show a very poor prediction in the model. In the maps, the cells with yellow colour show a better estimation in the OLS model. In general, the residuals distribute differently based on locations in the study area, which means the necessity to go to local regression, thus, the GWR model is used to explain the spatial heterogeneity.

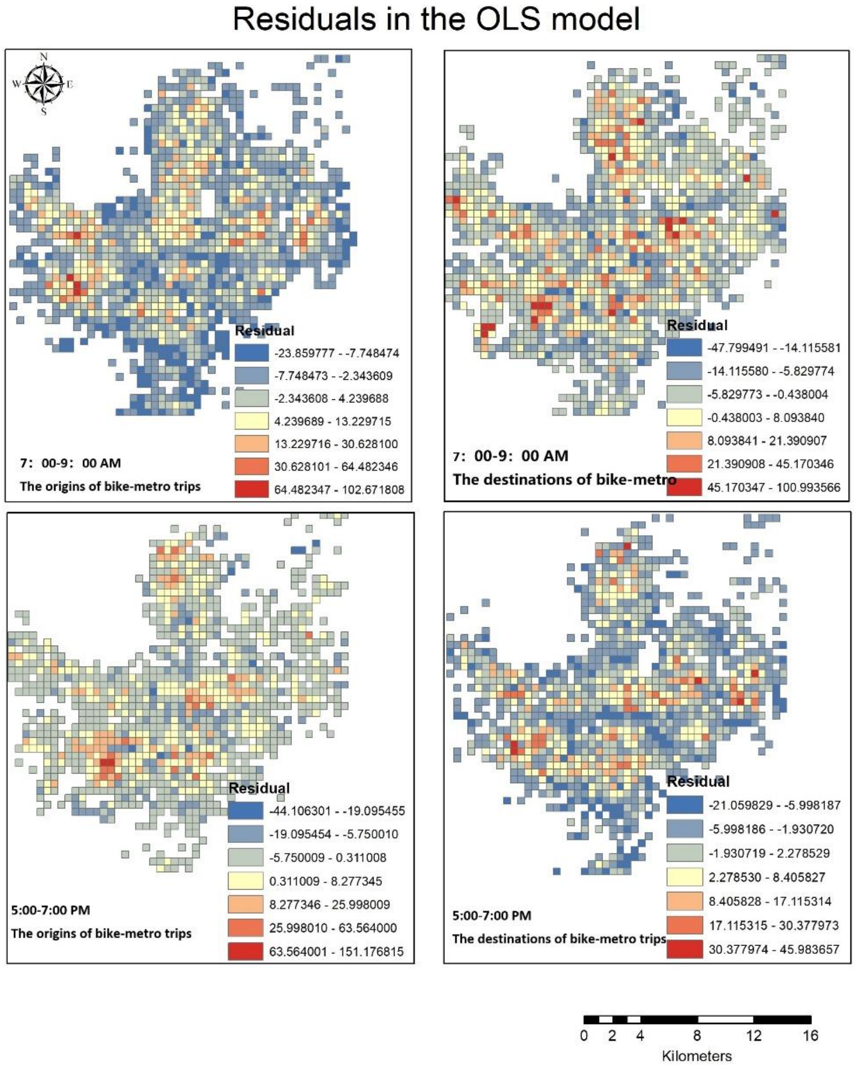


Figure 4-12: The distribution of residuals in the OLS model

4.3.2 Geographically weighted regression

4.3.2.1 GWR model analysis for the morning peak hours

(1) The number of origins as dependent variable

Combined with the method described in Chapter 3, the corresponding independent variables obtained by linear regression are input in the GWR4 software to model the distribution of the origins of shared bike trips during the morning peak in the study area. The system default golden section search method and AICc are used to search and determine the optimal bandwidth. The iterative steps show in the *Appendix 8*.

The statistical values of the parameter estimation in the GWR model are finally shown in *Table 4-9*, including: minimum, maximum, mean and standard deviation. First, the R-square indicates that the model can explain about 67% of the variance in the dependent variable. Second, from the minimum and maximum of the regression parameter estimates, the coefficients of the input variables are both positive and negative, indicating that these variables have both an **increasing** effect and a **suppressing** effect on the number of origins of the bike-metro trips in the morning peak. Furthermore, through the mean of the regression parameters of various variables, overall, the input variables related to the land use types of residential, commercial, recreation, and healthcare will promote the distribution of the origins of the shared bike trip. Besides, the standardized deviation is used to measure the reliability of the estimated value of each coefficient. A small standard error indicates a reliable estimation.

In addition, through the geographical variability test function in GWR4.0 software, the heterogeneity of each variable was tested. The test verifies the geographical variability of the coefficients of various variables through model comparison. The test results are shown in the *Table 4-10* below. The columns "F" and "DOF for F test" are the results of F-test and the its degrees of freedom, which basically show the process of the model comparison. The most important column is the "DIFF of Criterion", which shows the difference in model comparison indicator between the original GWR model and the switched GWR model. If the switched GWR model attains a statistically better fit, the value of the model comparison indicator is smaller than that of the original GWR model so that "Diff of Criterion" becomes a positive value, suggesting no spatial variability in the highlighted local term (GWR4 user manual, 2012). Therefore, there is basically no spatial heterogeneity in other variables except "**residence**".

Table 4-9: Independent coefficient distribution

Variables	Min.	Max.	Mean	Std. dev
Intercept	0.386892	10.358752	3.663497	1.717908
Residence	**-0.082634**	**0.492218**	**0.060450**	**0.070391**
Commercial	-0.208254	0.945139	0.212650	0.194403
Recreation	-1.015101	1.714205	0.117600	0.373027
Healthcare	-1.799641	2.940642	0.159692	0.336183
R^2	0.671852			
Adjusted R^2	0.583723			

Table 4-10: The geographical variability tests

Geographical variability tests of local coefficients
****

Variable	F	DOF for F test	DIFF of Criterion
Intercept	3.278702	34.933 1021.053	-33.330359
healthcare	1.921602	26.460 1021.053	14.074225
recreation	1.852844	27.939 1021.053	17.005409
shopping	1.743750	28.127 1021.053	20.588843
residence	2.722723	29.034 1021.053	**-10.524884**

-------------------- ------------------ ---------------- -----------------

Note: positive value of diff-Criterion (AICc, AIC, BIC/MDL or CV) suggests no spatial variability in terms of model selection criteria.
F test: in case of no spatial variability, the F statistics follows the F distribution of DOF for F test.

(2) The number of destinations as dependent variable

Combined with the above analysis, the same method is used to model the destinations of bike-metro trips during morning peak hours of working day in the study area. The result of the coefficients and geographical variability test is shown in *Table 4-11* and *Appendix 9*. From the results, there is a spatial heterogeneity in the coefficient estimates of variable "**job**", and "**education**". These variables have both an increasing effect and a suppressing effect on the number of destinations of the bike-metro trips in the morning peak. In regions with a large number of jobs, and educational places, the destinations of shared bike trips on the morning peak of weekdays is not evenly distributed.

Table 4-11: Independent coefficient distribution

Variables	Min.	Max.	Mean	Std. dev
Intercept	0.549801	12.176621	3.705771	1.826075
Job	**-0.290245**	**0.339193**	**0.089550**	**0.070937**
Commercial	-0.541084	0.433732	-0.090745	0.159952
Recreation	-0.772419	1.345137	0.098271	0.317831
Education	**-3.063364**	**4.602606**	**0.712649**	**1.320932**
Healthcare	-0.905575	0.937207	0.218722	0.247152
R^2	0.706787			
Adjusted R^2	0.617103			

4.3.2.2 GWR model analysis for the evening peak hours

(1) The number of origins as dependent variables

From the mean value of the coefficients *Table 4-12*, in most cases, the increase in the number of job POIs, recreation POIs, education POIs, and healthcare POIs will increase the distribution of start points of shared bike trips on weekday evening peaks. The increase of the commercial places count will slightly decrease the distribution of start points in the evening peak. The results of the geographical variability test in *Appendix 9* show that the variable of the **job** count has different impact on the shared bike use in different locations.

Table 4-12: Independent coefficient distribution

Variables	Min.	Max.	Mean	Std. dev
Intercept	-0.210181	11.231436	2.812781	1.716721
Job	**-0.087458**	**0.250871**	**0.049061**	**0.052444**
Recreation	-0.431996	1.153375	0.135687	0.288040
Education	-1.832886	3.705950	0.256193	0.968621
commercial	-0.393990	0.297615	-0.018294	0.089885
Healthcare	-0.961501	1.079855	0.096126	0.253023
R2	0.692069			
Adjusted R2	0.611648			

(2) The number of destinations as dependent variable

The same method is used to model the origins of bike-metro trips during evening peak hours of working day in the study area. The distribution of the model parameters obtained by the GWR analysis is shown in *Table 4-13*. From the mean of the coefficients, overall, all variables that related to job, residence, recreation, and healthcare have positive effects on the distribution of destinations of bike-metro trips. The result of geographical variability test are shows in *Appendix 9*. From the result, there is no spatial heterogeneity in the variables of the number of job POI, healthcare POI and government POI. Thus, they are converted into global variables. The impact of the number of "**residences**" and "**recreation**" on the shared bike use varies according to geographical locations. In some places, the number of residential and recreation areas may promote bike-metro trips, while in some places, the number of residential and recreational venues will inhibit bike-metro trips.

Table 4-13: Independent coefficient distribution

Variables	Min.	Max.	Mean	Std. dev
Intercept	-0.274906	13.999285	3.018856	1.865275
Job	-0.249677	0.188156	0.006941	0.050490
Residence	**-0.138351**	**0.425314**	**0.046195**	**0.067072**
Recreation	**-0.812463**	**1.698575**	**0.309736**	**0.390290**
Healthcare	-1.026656	1.129448	0.143665	0.237314
R2	0.687154			
Adjusted R2	0.580663			

4.3.2.3 Analysis and comparison of OLS and GWR model

The tables below (*Table 4-14* and *Table 4-15*) show the comparison of OLS model and the GWR model in detail. According to the values of indicators in the table, the AICc value in the GWR model results is lower than that in the global regression analysis, which means the model provides a better fit to the observed data (Esri, n.d.-a). Thus, the GWR model can better explain the data. The R-square and the adjusted R-square in the GWR model results are higher than those of the global regression analysis. This shows that in the GWR model, the independent variables have stronger explanatory power to the dependent variable and the model is more accurate. The residual squares can be used to judge the fitting effect of the model. The smaller this measure, the closer the fit of the GWR model to the observed data (Esri, n.d.-a). The residual squares in the GWR model results are smaller than those in the global regression analysis. Besides, the GWR model can reveal the different effects of each independent variable on the dependent variable in different spatial locations. Therefore, the GWR model is reliable in this study.

Table 4-14: Model comparison results (in the morning peak)

Morning peak				
Dependent variables	**The number of origins of "to metro station" trips**		**The number of destinations of "from metro station" trips**	
Indicators	OLS	GWR	OLS	GWR
AICc	7148.681880	6706.611851	7519.682936	7173.218864
R2	0.303000	0.671852	0.479000	0.706787
Adjusted R2	0.300000	0.583723	0.476000	0.617103
ResidualSquares	137603.990130	12126.616556	28125.599066	13895.381134

Table 4-15: Model comparison results (in the evening peak)

Evening peak				
Dependent variables	**The number of origins of "to metro station" trips**		**The number of destinations of "from metro station" trips**	
Indicators	OLS	GWR	OLS	GWR
AICc	5816.712838	5477.119255	5904.291943	5535.256845
R2	0.360000	0.692069	0.307	0.687154
Adjusted R2	0.356000	0.611648	0.304	0.580663
ResidualSquares	16622.350401	7593.292814	17548.72938	8061.805412

4.3.2.4 The interpretation of the GWR results

From the results of GWR model, in different locations, different factors play a role in explaining the distribution of bike-metro trips. This section will combine the conclusions drawn by the GWR model in this chapter to focus on the visualization of factors with spatial heterogeneity.

(1) In the morning peak hours

- **The number of origins as dependent variables**

The influence of the number of residences on the distribution of the origins of shared bike trips has spatial heterogeneity. In other words, the degree of impact of the number of residences on the use of shared bikes varies depending on the geographic locations. Therefore, visualize the spatial distribution of the coefficient values of the "**residence**" variable obtained from the GWR analysis.

From *Figure 4-13*, in the areas with **red** colour, there is a good link between the number of residential units and the use of bikes, while the areas with **blue** colour show a poor relationship between the residential units and the bike-metro trips. Specifically, the number of residences has an obvious positive effect in the southwest of the city near the end station of Metro Line 3 (the black box in the figure), and it has almost no positive impact on the north, south and east of the city, and even has an inhibitory effect in the north, south and east of the city(the red box).

Among them, the southwestern area of the city is the **High-tech Development Zone** of Xi'an City, which has a relatively high degree of development. There are densely distributed residences near the metro terminal station. Besides, the density of public transport network in this area is relatively low. Thus, the demand for shared bikes in this area is high, and the region will generate more shared bike trips that relate to commuting during the morning peak. On the other hand, there are more commercial districts, enterprises, and schools in the south and north of the city. Therefore, there are relatively few shared bike trips from these areas to metro stations in the morning peak. On the east side of the city, the coefficient of the "the number of residence POIs" is small, because the development intensity here is relatively low and there is generally little demand for shared bike use.

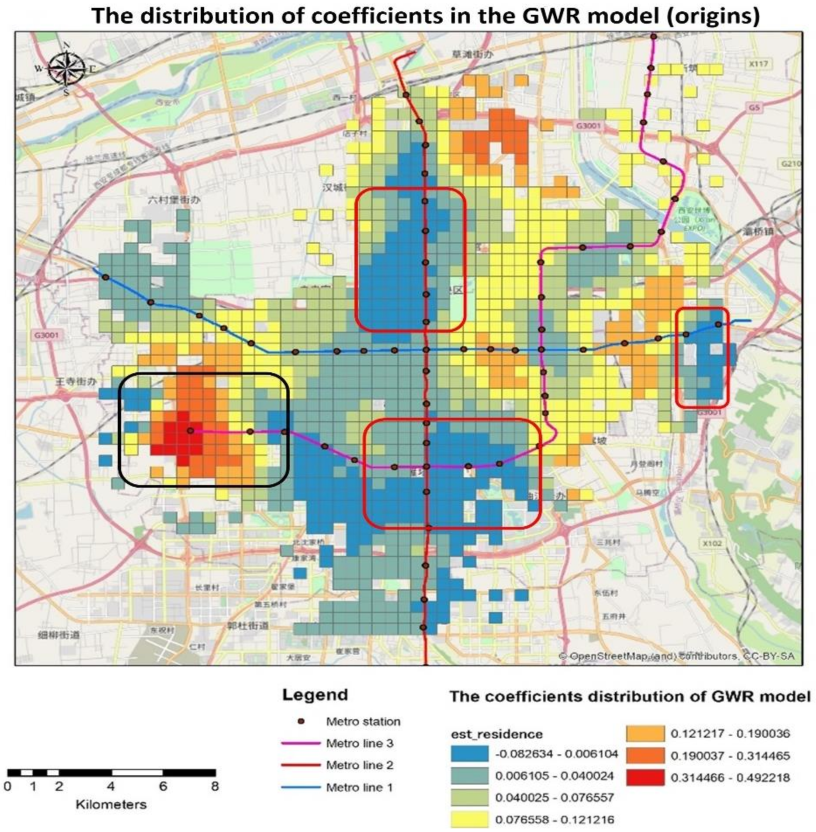


Figure 4-13: The distribution of the coefficients of variable "the count of residence POIs"

➢ **The number of destinations as dependent variables**

From the geographical variability test in the GWR model, the degree of impact of variables "job" and "education" on the use of shared bikes varies depending on the geographic locations. Therefore, the maps below visualize the spatial distribution of the coefficient values of "**job**" and "**education**". From the first map (*Figure 4-14*), the red area represents the high value of the coefficient, which means the number of jobs has a significant positive effect in the southeast of the city near the Metro Line 3. The blue areas with a negative value of coefficient indicate the distribution of job locations has an inhibitory effect on shared-bike trips in the northeast area.

Among them, in the southeast of the city (the black box in the map), there is a concentration of **cultural media companies, and some scientific research institutions**. Thus, the number of job locations in this area will greatly influence the shared bike trips in the morning peak hours. As discussed above, the southern part of the city is mainly commercial land and has not yet opened in the morning. Therefore, it is less attractive to shared bike trips in the morning peak. The northeast part of the city is mainly green land, and some new campuses of schools. Therefore, the cycling activities have nothing to do with the "job", and even slightly decrease with the increase of the jobs.

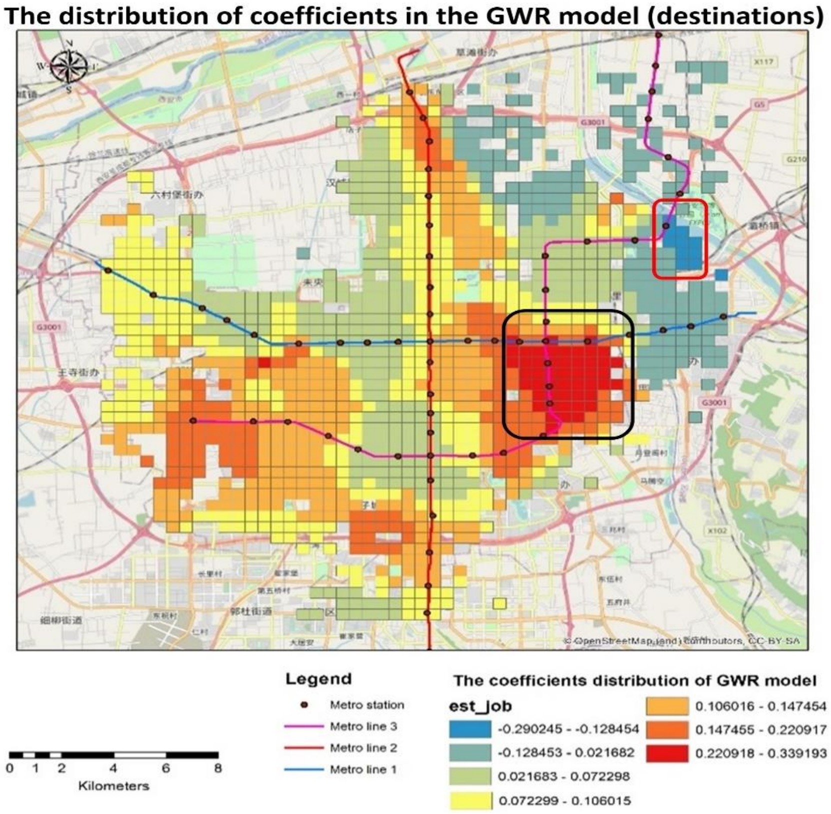


Figure 4-14: The distribution of the coefficients of variable "the count of job POIs"

As for the coefficient distribution of the variable "**education**", the number of educational places has a positive effect in the western part of the city and part of the northeast, as well as the area close to Hujiamiao and Shijiajie Station (*Figure 4-15*). These places are mainly residential areas, and there are many primary education institutions, such as elementary schools, junior high schools and kindergartens. Most of these schools are located at the end of metro lines, far from the city center, and mostly serve residents living in social housing (Chen, 2015). The coverage of public transportation in these areas is low. Thus, there is potentially a greater demand for shared bikes. However, in the region close to Shijiajie Station and Hujiamiao Station, the public transport is convenient, the number of education POIs still has a higher degree of influence in this area, which indicating that there are generally more education-related cycling activities during the morning peak on weekdays.

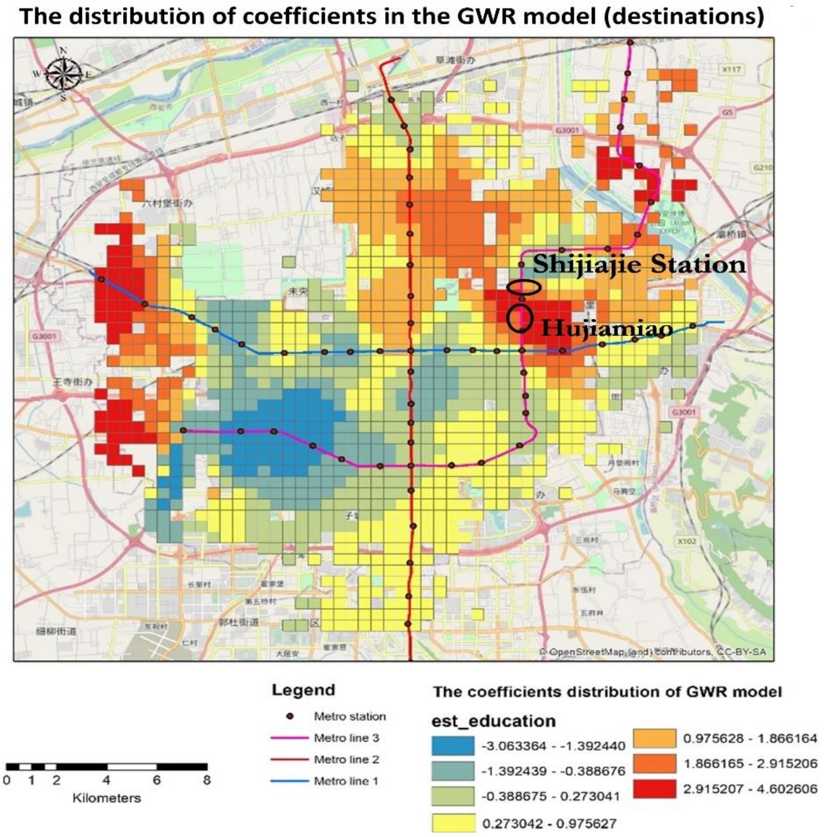


Figure 4-15: The distribution of the coefficients of variable "the count of education POIs"

(2) In the evening peak hours

➢ **The number of origins as dependent variables**

From the geographical variability test in the GWR model, the degree of impact of variables "**job**" on the use of shared bikes varies depending on the geographic locations. Therefore, the map below visualized the spatial distribution of the coefficient values of the number of jobs.

From *Figure 4-16*, the distribution of job locations near the transfer station of Metro Line 1 and Line 2 and the transfer station of Line 1 and Line 3 has a clear positive effect on the distribution of the origins of shared bike trips. The transfer stations often attract large numbers of passengers to gather around in the commuting peak hours, which is a good explanation for the greater the number of job locations in the region, the more attracted bike rides from these places to the transfer stations. In the high-tech development zone, the distribution of companies and enterprises generally has a positive impact on the distribution of the origins of shared bike trips. Because there are a large number of commuters who depart from the job locations and go to the metro stations in the evening peak hours.

To the east of the city, the negative coefficients indicate the distribution of origins for shared bike trips slightly decreases as the number of job locations increases, because the area is mainly green space. In the evening rush hour, there is less demand for shared bikes in the area, thus, the bike trips rarely related to the distribution of job locations.

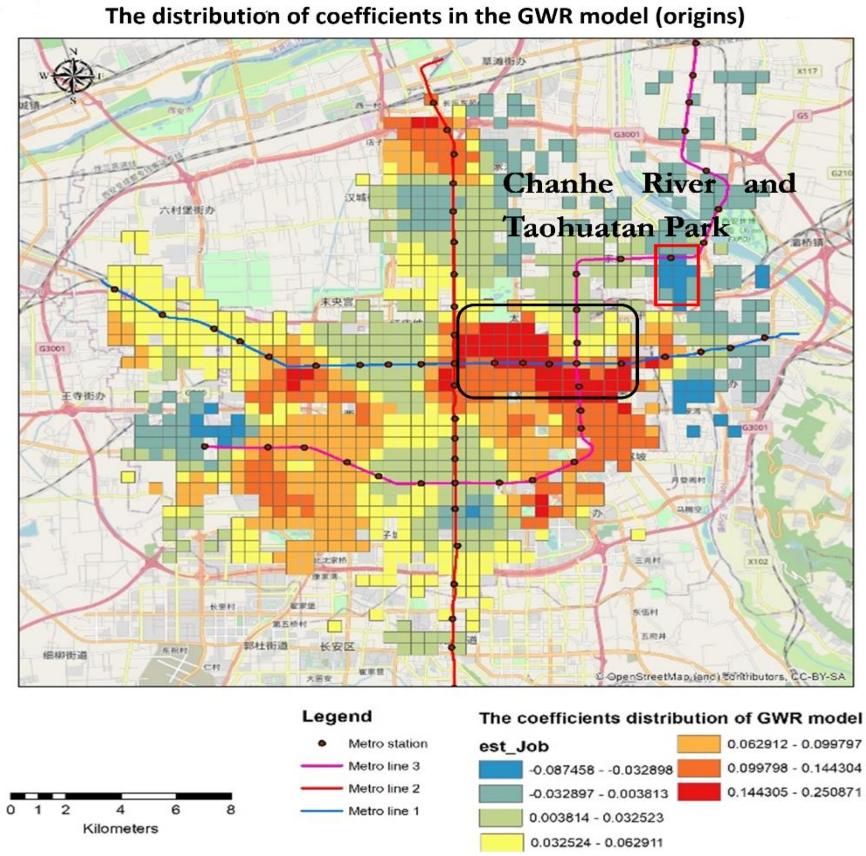


Figure 4-16: The distribution of the coefficients of variable "the count of job POIs"

➢ **The number of destinations as dependent variables**

The results of GWR analysis show that the number of **residence** and **recreation** has a spatial heterogeneity on the impact of the distribution of the bike trips destinations in the evening peak. Therefore, the maps below (*Figure 4-17,4-18*) visualize the spatial distribution of the estimated parameter values of the number of residences, and recreation.

Similar to the impact of the number of residences on the distribution of shared bike trips during the morning peak hours, the number of residences near the terminal station of Metro Line 3 in the high-tech development zone has an obvious positive effect on the distribution of the destinations of shared bike trips. Because of the sparse bus lines in this dense residential area, it will attract a large number of people ride home from metro stations. It can be seen from *Figure 4-17*, the number of recreation venues has the most significant positive influence on the distribution of cycling destinations in the area near the Zhangbabeilu Station in the high-tech zone. There are also some popular recreation places in the region, the bike trips related to recreation are very common in this area during evening peak.

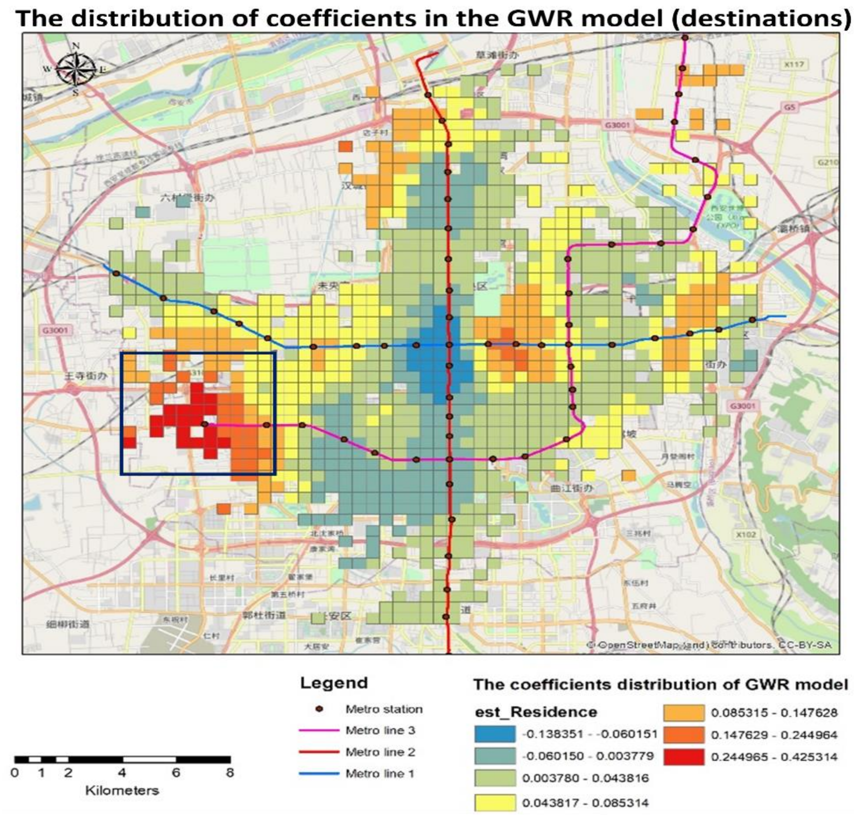


Figure 4-17: The distribution of the coefficients of variable "the count of residence POIs"

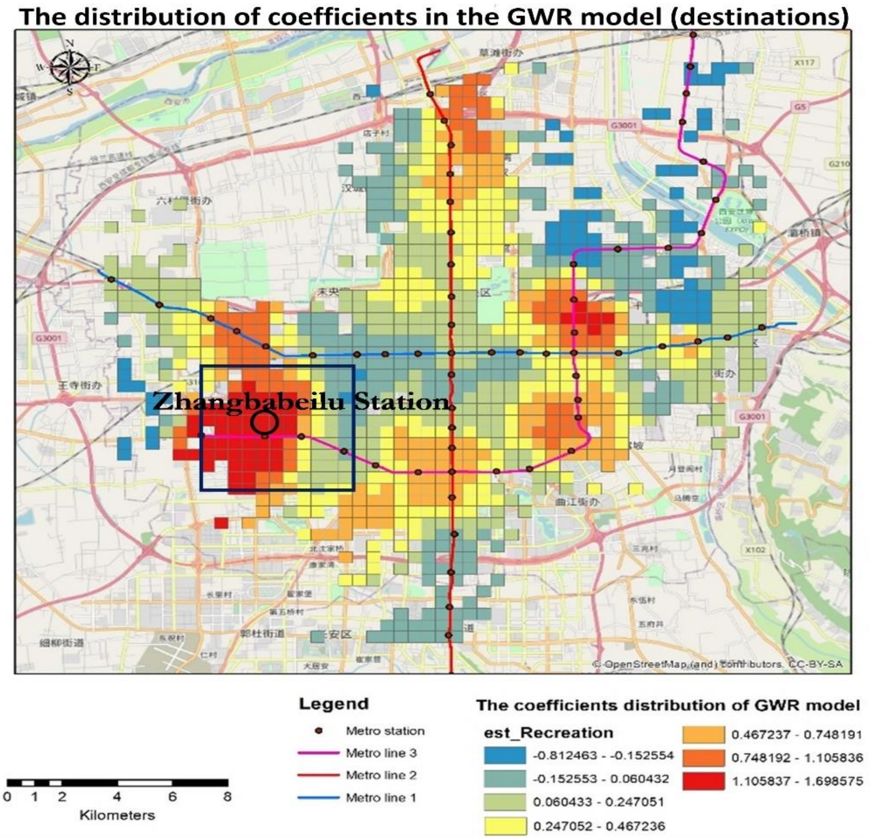


Figure 4-18: The distribution of the coefficients of variable "the count of recreation POIs"

4.4 Disussion

4.4.1 The spatiotemporal characteristics of shared bike use for metro system through visualization

In this study, the spatiotemporal distribution characteristics of the shared bike use as access and egress of metro stations in Xi'an city were revealed. From the **temporal** perspective, the research is mainly looking at the usage variation of the selected workday. The use of shared bikes during the day appears to have two peaks from 7:00 to 9:00 in the morning and from 5:00-7:00 in the evening, and the number of trips in these two periods is significantly greater than in other periods. Among them, the bike-metro trips accounted for more than 20% of all shared bike trips during the morning rush hour. These findings are similar to the studies by Yang (2019), which focused on Nanchang city. However, the research of Li (2019), and Yang (2019) shows that the variation in the usage of shared bikes during the workdays may occur in a single peak period or no peak in Nanjing city. This means that different size, topography, and economic development level of cities generally has different use patterns of the shared bikes (Du, 2019).

On the other hand, from the **spatial** perspective, different visualization techniques combined with data processing methods are primarily used to identify the spatial characteristics of shared bike use for metro in the study area. According to the findings on the distribution of trip **O-Ds**, the distribution of O-Ds is quite different in the morning and evening peak hours. Moreover, there are obvious hot spots in the **city center**, near the **metro transfer station**, and in the **high-tech zone**. Due to its high degree of development, the city center is an attractive place for companies and enterprises, it also concentrates commercial and educational facilities leading to an overall high travel demand. These findings are consistent with other studies, such as Du (2019), and El-Assi (2017). In addition, results are also consistent with the result of other studies that use Xi'an as a case city (Cheng, 2019). The conclusions of this study on the distribution of **cycling paths** and the distribution of cycling O-Ds are basically consistent. Compared with the network-based method discussed in other related studies, the visualization method used to generate riding paths is closer to the actual road network distribution. However, the "route plan" API returns the **optimal routes** with the shortest time or the shortest riding distance, hence, to a certain extent, the users' actual choice behaviour and actual road conditions were ignored.

4.4.2 Built environment effects on bike-metro trips

Due to the limitation of data availability, this study mainly discussed the influencing factors related to urban land use based on POI data. The method consists of using the number of different POIs in a certain range as different independent variables to explore its impact on the use of shared bikes, an similar approach also seen in the studies of Ma, Zhang, Li, Wang and Zhao (2019), and Ma, Cao, and Jin (2019). However, POI data represents residential buildings, office buildings, factories, schools, etc. in the form of points. The residential buildings, office buildings or shopping malls often contain many different households, offices, or shops that can show stronger attraction than only a point (Moridpour, Toran Pour, & Saghapour, 2019). In short, representing these places by points will cause a deviation in quantity. This is also one of the reasons why the coefficient estimates in GWR results are generally low.

On the other hand, to reveal the link between these POI-based independent variables with the number of O-Ds of bike-metro trips, the GWR model was used to investigate the land use-related variables considering the spatial heterogeneity. The idea is consistent with the studies of Ma (2019), and Cardozo, García-Palomares, and Gutiérrez (2012). Inspired by the research of Yang (2019), and Xu (2019), the analysis of the influencing factors is based on 500 meter*500 meter cells. In order to make the research results more accurate, this study also experimented with 200m * 200m and 100m * 100m grids for global regression testing. However, the fitting degree did not improve. The proportion of each type of POI in the grid was also used as an independent variable for the study, but it also did not significantly improve the model fit since the R^2 is basically around 0.2, which means the model can only explain 20% of the variation of the dependent variable. Therefore, the final determination was to use the 500-meter grid as the unit and the number of each type of POI as the independent variable in the research. The global regression results and the final results of the GWR model are comparable with the study of Cheng (2019) who uses a similar dataset and the same case study area.

4.4.3 Recommendations for improving the shared bike system

The areas near Kejilu, Yanpingmen and Zhangbabeilu Station in the **High-tech Zone** are concentrations of jobs. From the results, these areas are not only hot spots for the distribution of the origins and destinations of the bike-metro trips, but are also greatly influenced by the number of companies. Therefore, combining the **challenges** caused by the booming of DLBS bikes that discussed before, it is necessary to **replenish the bikes in time** in these hot spots during morning and evening peak hours to meet the needs of commuters who use shared bikes to connect with the metro system. In other words, it is important to ensure the efficiency of the **rebalance** in these areas.

It is worth mentioning that in some areas, the coefficient value of the GWR model is high, but it is not an obvious hot spot for bike use from the visualization results. This means some specific land-use factors play a very important role in some areas that did not show an obvious concentration of bike-metro trips. For example, in the southeast area along Metro Line 3, and the transfer station of Line 2 and Line 3, the number of jobs has the most significant positive influence on the shared bike use. However, from the visualization results, during the morning rush hours, the distribution of the destinations is not obviously clustering in this area. Because the area is close to the city center with diverse land use, the road network is dense, and public transportation is convenient. Although the area can attract more work-related shared bike trips during the morning rush hours, the area is not dominated by companies, so the shared bike trips did not concentrate during the morning rush hours. In other words, the use of shared bikes in these areas is sensitive to a certain type of land use. Therefore, with the expansion of the job locations, it is necessary to supplement the number of shared bikes in this area in time and improve the supporting facilities such as the **designated parking areas** to guide users to **park the bikes**.

In addition, in some hot spot areas, the coefficient of influencing factors has a minimum value (such as the city center, the location of the high-tech zone near the city center), which often indicates that these areas have high land use diversity, and are slightly influenced by a certain of land use. The use of shared bikes in these areas generally does not change significantly due to the increase of certain types of land use. Therefore, more attention should be paid to the **maintenance** of the bikes and corresponding bike lanes in the area to ensure the efficiency of shared bike use. Besides, considering the usage of the shared bikes in these areas, some designated parking spaces can be planned to deal with parking disorder.

From the perspective of **cycling routes distribution** (section 4.2.3.2), there are dense road network in the central area of the city, but none of them is the most frequently used route for cycling. Because the development intensity of the downtown area is the highest and the land use is diverse. The cycling activities in this area have no clear purpose. Therefore, the riding routes are also scattered. The hot spots for cycling routes are mainly distributed in the **high-tech zone**, **the administrative center** on the north side of the city, and the areas near the stations along Line 3 on the east side of the city,which are high-densely job or residential areas. Because the origins or destinations of these paths are mostly related to commuting, such as residential or job-related places. In addition, as can be seen from the road network distribution map of Xi'an (3.2.2.3), the road network in these areas is less dense than in the city center, so it is more likely to generate some high-frequency road sections. However, this also means a risk of mixed traffic on bikes and motorised vehicles. Thus, in these road sections, it is necessary to ensure the safety of riding environment. The planners should pay attention to increase the **coverage of bike lanes**, and to **maintain** the existing bike lanes in this area. The table below (*Table 4-16*) summarized the result combintions of the visualization and the GWR model, and the relevant suggestions for planners.

Table 4-16: Different combinations of the visualization and GWR model results

Visualization / GWR	High (Hot spot)	Low (Non-hotspot)
High coefficients	These areas are often rich in POI types, and there are a lot of commuting trips in the morning and evening peaks. For the system improvement, the vehicles need to maintain and rebalance in time.	Morning and evening peak cycling activities are not clearly gathered at this location, but there are bike-metro trips related to certain specific activities (such as commuting). It is necessary to pay attention to the expansion of land use and supplement of shared bike-related facilities.
Low coefficients	The land use diversity is on a high level and the demand for shared bikes is large, but there is no obvious cycling activity related to a certain type of land use in the morning and evening peak. It is necessary to ensure the maintenance of facilities related to shared bikes (such as parking area, bicycle lanes, etc.) in this area.	Generally appears in areas with low development intensity. It is important to add vehicles in time in these places according to the planning related to land use.

5 CONCLUSION

To achieve **objective 1**, and **objective 2**, which is to provide insight in the spatiotemporal pattern of dockless shared bike usage for trips to/from metro stations, the GPS point data of tens of millions of shared bike trips was analysed in the Python environment. The O-D information of the bike trips was extracted. A method for identifying shared bike trips for connecting to the metro system was proposed. In order to visualize the spatial and temporal distribution of shared bikes from different aspects, this research also proposed a method of integrating O-D, and a method of generating cycling paths based on Route Plan API on Baidu Map Server. With the data preparation, different visualization tools such as *statistical graphs*, *heat maps*, *interactive flow maps*, and *route distribution maps* are used to show the distribution patterns of shared bike trips related to the metro.

The analysis showed that in the study area, most of the shared bike trips are taken in the morning peak and have the most frequent cycling activities related to the metro. Besides, in the morning peak, there are a lot of cycling activities starting from the **residential areas**, and ending a **metro station**, or starting from the **metro station**, riding to the **work place**. The total amount of cycling activities in the evening peak is not as much as that in the morning peak, but the hot spots of cycling distribution have a reverse trend with the morning peak. This indicates that riding activities related to the metro are largely related to *commuting*. This is a pattern that should lead planners to focus on the management of **parking spaces**， **rebalance** of bikes in the residential and company intensive areas during peak hours.

To achieve **objective 3**, which is to identify factors that influence the use patterns of bike-sharing for connecting with the metro stations, the POI data obtained from Baidu map server is used as a source of factors related to the distribution of land use. In this way, the relationship between the land use and the distribution of the origins and destinations of bike-metro trips in morning and evening peak hours were investigated. Consistent with the visualization results, in the global regression, *residence* count has the greatest impact on the cycling origins. Commuting-related variables such as the number of *job and educational POIs* have the greatest impact on the cycling destinations **in the morning peak**; **In the evening rush hour**, *work-related* variables have the greatest impact on the cycling origins, and the number of *residential and recreation* POIs have the greatest impact on the cycling destinations. In addition, the GWR reveals that during the two peak periods, residential and job distribution have spatial heterogeneity on the impact of the origins and destinations of bike-metro trips respectively. This means that the influence of these types of land use on the bike-metro trips varies from different regions in the study area. In conclusion, the distribution of cycling hotspots depends largely on the type of land use that related to commuting activities.

Limitation and recommendation for future work

The study used data processing methods combined with different visualization techniques to reveal the shared bike use variation in selected workdays, especially during peak hours. The GWR model was used to explain the link between the shared bike use for metro and the land use related factors. In this way, the research method can be applied after adaptations to understand the shared bike use as access and egress for metro system in other contexts from the spatial and temporal perspectives, as well as the influencing factors so as to solve the first and last mile problem more effectively.

First, this study conducts research on the basis of available shared bike data for three workdays. As the usage of the selected three days was found to be very similar in the study, the data of one day was selected for detailed analysis. Due to the volume of the data and computation time, a longer time frame was not into consideration. In the future, one week or two weeks data can be used for a more comprehensive analysis that compare the variation on workdays and weekends.

Second, some socio-demographic and transport-facility factors can be considered if data is available. In this way, the influencing factors can be considered more comprehensively and the results are more convincing.

Third, in line with the discussion, there are some drawbacks using POI data. In future research, it is necessary to use more advanced methods or more accurate indicators to reduce the limitations of POI data.

LIST OF REFERENCE

Alam, S. (2010). Factors in deciding Metro Rail in Developing Countries : A study on the proposed Metro Rail system for Dhaka, *3*(December), 33–51.

Alcorn, L. G., & Jiao, J. (2019). Bike-Sharing Station Usage and the Surrounding Built Environments in Major Texas Cities. *Journal of Planning Education and Research*, (April). https://doi.org/10.1177/0739456X19862854

Ashqar, H. I., Elhenawy, M., Almannaa, M. H., Ghanem, A., Rakha, H. A., & House, L. (2017). Modeling bike availability in a bike-sharing system using machine learning. *5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2017 - Proceedings*, 374–378. https://doi.org/10.1109/MTITS.2017.8005700

Ashqar, H. I., Elhenawy, M., & Rakha, H. A. (2019). Modeling bike counts in a bike-sharing system considering the effect of weather conditions. *Case Studies on Transport Policy*, *7*(2), 261–268. https://doi.org/10.1016/j.cstp.2019.02.011

Baidu. (n.d.-a). androidsdk | 百度地图API SDK. Retrieved June 12, 2020, from https://lbsyun.baidu.com/index.php?title=androidsdk/guide/search/poi

Baidu. (n.d.-b). Web服务API | 百度地图API SDK. Retrieved June 11, 2020, from http://lbsyun.baidu.com/index.php?title=webapi/direction-api-v2

Bao, J., He, T., Ruan, S., Li, Y., & Zheng, Y. (2017). Planning bike lanes based on sharing-bikes' trajectories. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *Part F1296*, 1377–1386. https://doi.org/10.1145/3097983.3098056

Bao, J., Liu, P., Yu, H., & Xu, C. (2017). Incorporating twitter-based human activity information in spatial analysis of crashes in urban areas. *Accident Analysis and Prevention*, *106*(July), 358–369. https://doi.org/10.1016/j.aap.2017.06.012

Bao, J., Shi, X., & Zhang, H. (2018). Spatial Analysis of Bikeshare Ridership With Smart Card and POI Data Using Geographically Weighted Regression Method. *IEEE Access*, *6*, 76049–76059. https://doi.org/10.1109/ACCESS.2018.2883462

Barbour, W., Majewski, A., Shoup, L., Kopstain, E., Philip, C., & Work, D. B. (2019). Data-driven methods for dockless bike infrastructure planning. *2019 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2019*, 578–583. https://doi.org/10.1109/PERCOMW.2019.8730590

Bidanset, P., & Lombard, J. (2014). The Effect of Kernel and Bandwidth Specification in Geographically Weighted Regression Models on the Accuracy and Uniformity of Mass Real Estate Appraisal. *Journal of Property Tax Assessment & Administration*, *10*(3), 5–14.

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

Canitez, F. (2019). A socio-technical transition framework for introducing cycling in developing megacities: The case of Istanbul. *Cities*, *94*(May), 172–185. https://doi.org/10.1016/j.cities.2019.06.006

Cardozo, O. D., García-Palomares, J. C., & Gutiérrez, J. (2012). Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. *Applied Geography*, *34*, 548–558. https://doi.org/10.1016/j.apgeog.2012.01.005

Caulfield, B., O'Mahony, M., Brazil, W., & Weldon, P. (2017). Examining usage patterns of a bike-sharing scheme in a medium sized city. *Transportation Research Part A: Policy and Practice*, *100*, 152–161. https://doi.org/10.1016/j.tra.2017.04.023

Chen, R. (2019). "Bike litter" and obligations of the platform operators: Lessons from China's dockless sharing bikes. *Computer Law and Security Review*, *35*(5), 105317. https://doi.org/10.1016/j.clsr.2019.03.011

Chen, Z., van Lierop, D., & Ettema, D. (2020). Dockless bike-sharing systems: what are the implications? *Transport Reviews*, *0*(0), 1–21. https://doi.org/10.1080/01441647.2019.1710306

Cheng. (2019). 基于GWR的共享单车出行特征及影响因素空间异质性研究. Retrieved May 17, 2020, from http://gb.oversea.cnki.net/KCMS/detail/detail.aspx?filename=1019628294.nh&dbcode=CMFD&dbname=CMFDREF

Central Government of the People's Republic of China. (n.d.). 交通运输部解读《关于鼓励和规范互联网租赁自行车发展的指导意见》_部门政务_中国政府网. Retrieved October 17, 2019, from http://www.gov.cn/xinwen/2017-08/04/content_5215971.htm

Cho, S. H., Lambert, D. M., & Chen, Z. (2010). Geographically weighted regression bandwidth selection and spatial autocorrelation: An empirical example using Chinese agriculture data. *Applied Economics Letters*, *17*(8), 767–772. https://doi.org/10.1080/13504850802314452

Cole-Hunter, T., Donaire-Gonzalez, D., Curto, A., Ambros, A., Valentin, A., Garcia-Aymerich, J., … Nieuwenhuijsen, M. (2015). Objective correlates and determinants of bicycle commuting propensity in an urban environment. *Transportation Research Part D: Transport and Environment*, *40*(2), 132–143. https://doi.org/10.1016/j.trd.2015.07.004

Columbia. (n.d.). Geographically Weighted Regression | Columbia Public Health. Retrieved May 22, 2020, from https://www.publichealth.columbia.edu/research/population-health-methods/geographically-weighted-regression

Corcoran, J., Li, T., Rohde, D., Charles-Edwards, E., & Mateo-Babiano, D. (2014). Spatio-temporal patterns of a Public Bicycle Sharing Program: The effect of weather and calendar events. *Journal of Transport Geography*, *41*, 292–305. https://doi.org/10.1016/j.jtrangeo.2014.09.003

de Souza, F., La Paix Puello, L., Brussel, M., Orrico, R., & van Maarseveen, M. (2017). Modelling the potential for cycling in access trips to bus, train and metro in Rio de Janeiro. *Transportation Research Part D: Transport and Environment*, *56*, 55–67. https://doi.org/10.1016/j.trd.2017.07.007

DeMaio, P. (2009). Bike-sharing: History, Impacts, Models of Provision, and Future. *Journal of Public Transportation*, *12*(4), 41–56. https://doi.org/10.5038/2375-0901.12.4.3

DeMaio, P., & Gifford, J. (2004). Will Smart Bikes Succeed as Public Transportation in the United States? *Journal of Public Transportation*, *7*(2), 1–15. https://doi.org/10.5038/2375-0901.7.2.1

DSS. (n.d.). DSS - Interpreting Regression Output. Retrieved June 11, 2020, from https://dss.princeton.edu/online_help/analysis/interpreting_regression.htm

Du, M., & Cheng, L. (2018). Better understanding the characteristics and influential factors of different travel patterns in free-floating bike sharing: Evidence from Nanjing, China. *Sustainability (Switzerland)*, *10*(4). https://doi.org/10.3390/su10041244

Du, M., Cheng, L., Li, X., & Yang, J. (2019). Investigating the Influential Factors of Shared Travel Behavior: Comparison between App-Based Third Taxi Service and Free-Floating Bike Sharing in Nanjing, China. *Sustainability*, *11*(16), 4318. https://doi.org/10.3390/su11164318

Du, Y., Deng, F., & Liao, F. (2019). A model framework for discovering the spatio-temporal usage patterns of public free-floating bike-sharing system. *Transportation Research Part C: Emerging Technologies*, *103*(May 2018), 39–55. https://doi.org/10.1016/j.trc.2019.04.006

El-Assi, W., Salah Mahmoud, M., & Nurul Habib, K. (2017). Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. *Transportation*, *44*(3), 589–613. https://doi.org/10.1007/s11116-015-9669-z

Esri. (n.d.-a). Interpreting GWR results. Retrieved May 27, 2020, from http://resources.esri.com/help/9.3/arcgisdesktop/com/gp_toolref/spatial_statistics_tools/interpreting_gwr_results.htm

Esri. (n.d.-b). Regression analysis basics. Retrieved May 14, 2020, from http://resources.esri.com/help/9.3/arcgisdesktop/com/gp_toolref/spatial_statistics_toolbox/regression_analysis_basics.htm

Esri. (n.d.-c). Spatial Autocorrelation (Morans I) (Spatial Statistics). Retrieved June 11, 2020, from http://resources.esri.com/help/9.3/arcgisengine/java/gp_toolref/spatial_statistics_tools/spatial_autocorrelation_morans_i_spatial_statistics_.htm

Ester, M., Kriegel, H. P., & Sander, J. (1999). *Knowledge discovery in spatial databases. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 1701). https://doi.org/10.1007/978-3-642-60243-6_1

Etienne, C., & Latifa, O. (2014). Model-based count series clustering for bike sharing system usage mining: A case study with the vélib' system of Paris. *ACM Transactions on Intelligent Systems and Technology*, *5*(3), 1–21. https://doi.org/10.1145/2560188

Faghih-Imani, A., & Eluru, N. (2015). Analysing bicycle-sharing system user destination choice preferences: Chicago's Divvy system. *Journal of Transport Geography*, *44*, 53–64. https://doi.org/10.1016/j.jtrangeo.2015.03.005

Faghih-Imani, A., Hampshire, R., Marla, L., & Eluru, N. (2017). An empirical analysis of bike sharing usage and rebalancing: Evidence from Barcelona and Seville. *Transportation Research Part A: Policy and Practice*, *97*, 177–191. https://doi.org/10.1016/j.tra.2016.12.007

González, F., Melo-Riquelme, C., & de Grange, L. (2016). A combined destination and route choice model for a bicycle sharing system. *Transportation*, *43*(3), 407–423. https://doi.org/10.1007/s11116-015-9581-6

Xi'an municipal Government. (n.d.). Xi'an China. Retrieved October 15, 2019, from http://en.xa.gov.cn/

Griffin, G. P., & Sener, I. N. (2016). Planning for bike share connectivity to rail transit. *Journal of Public Transportation*, *19*(2), 1–22. https://doi.org/10.5038/2375-0901.19.2.1

Gu, T., Kim, I., & Currie, G. (2019). Measuring immediate impacts of a new mass transit system on an existing bike-share system in China. *Transportation Research Part A: Policy and Practice*, *124*(February), 20–39. https://doi.org/10.1016/j.tra.2019.03.003

Guo, Y., Zhou, J., Wu, Y., & Li, Z. (2017). Identifying the factors affecting bike-sharing usage and degree of satisfaction in Ningbo, China. *PLoS ONE*, *12*(9), 1–19. https://doi.org/10.1371/journal.pone.0185100

Haining, R. P. (2009). Spatial autocorrelation and the quantitative revolution. *Geographical Analysis*, *41*(4), 364–374. https://doi.org/10.1111/j.1538-4632.2009.00763.x

Hall, W. (2012). Built Environment and Demographic Predictors of Bicycle Access.

IXi'an. (n.d.). 在西安野蛮生长的51万辆共享单车，戴上了"紧箍". Retrieved March 2, 2020, from https://www.sohu.com/a/339041017_348945

J.Q. Wang, T.Cheng, J. H. (n.d.). Advances in Geo-Spatial Information Science - Google Books. Retrieved May 23, 2020, from https://books.google.nl/books?id=CAi6FKV00SYC&pg=PA26&lpg=PA26&dq=Fotheringham+bi-square&source=bl&ots=qXERGgJfoM&sig=ACfU3U1YsIJHQ92U_iO752juXZeCWwo3lQ&hl=en&sa=X&ved=2ahUKEwjzma7ltsrpAhXKDOwKHZafD-EQ6AEwAHoECAkQAQ#v=onepage&q=Fotheringham bi-square&f=fa

Jakaitiene, A. (2018). Nonlinear regression models. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, *1–3*, 731–737. https://doi.org/10.1016/B978-0-12-809633-8.20361-0

Ji, Y., Ma, X., Yang, M., Jin, Y., & Gao, L. (2018). Exploring spatially varying influences on metro-bikeshare transfer: A geographically weighted poisson regression approach. *Sustainability (Switzerland)*, *10*(5). https://doi.org/10.3390/su10051526

Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., & Banchs, R. (2010). Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing*, *6*(4), 455–466. https://doi.org/10.1016/j.pmcj.2010.07.002

Khoi, D. D., & Murayama, Y. (2012). *Multi-layer perceptron neural networks in geospatial analysis. Progress in Geospatial Analysis* (Vol. 9784431540). https://doi.org/10.1007/978-4-431-54000-7_9

Krizek, K. J., & Stonebraker, E. W. (2010). Bicycling and transit: A marriage unrealized. *Transportation Research Record*, (2144), 161–167. https://doi.org/10.3141/2144-18

L.Chen. (2015). 西部地区服务业集聚与城镇化互动发展研究. 中国社会科学出版社. Retrieved from https://books.google.nl/books?id=aqF8DwAAQBAJ

Lahoorpoor, B., Faroqi, H., Sadeghi-Niaraki, A., & Choi, S. M. (2019). Spatial cluster-based model for static rebalancing bike sharing problem. *Sustainability (Switzerland)*, *11*(11), 1–21. https://doi.org/10.3390/su11113205

Levy, N., Golani, C., & Ben-Elia, E. (2019). An exploratory study of spatial patterns of cycling in Tel Aviv using passively generated bike-sharing data. *Journal of Transport Geography*, *76*(January 2017), 325–334. https://doi.org/10.1016/j.jtrangeo.2017.10.005

Li, X., Zhang, Y., Sun, L., & Liu, Q. (2018). Free-floating bike sharing in jiangsu: Users' behaviors and influencing factors. *Energies*, *11*(7), 1–18. https://doi.org/10.3390/en11071664

Li, Y., Zhu, Z., & Guo, X. (2019). Operating characteristics of dockless bike-sharing systems near metro stations: Case study in Nanjing City, China. *Sustainability (Switzerland)*, *11*(8). https://doi.org/10.3390/su11082256

Lin, D., Zhang, Y., Zhu, R., & Meng, L. (2019). The analysis of catchment areas of metro stations using trajectory data generated by dockless shared bikes. *Sustainable Cities and Society*, *49*(January). https://doi.org/10.1016/j.scs.2019.101598

Lin, J. J., Zhao, P., Takada, K., Li, S., Yai, T., & Chen, C. H. (2018). Built environment and public bike usage for metro access: A comparison of neighborhoods in Beijing, Taipei, and Tokyo. *Transportation Research Part D: Transport and Environment*, *63*(1), 209–221. https://doi.org/10.1016/j.trd.2018.05.007

Liu, H. C., & Lin, J. J. (2019). Associations of built environments with spatiotemporal patterns of public bicycle use. *Journal of Transport Geography*, *74*(1), 299–312. https://doi.org/10.1016/j.jtrangeo.2018.12.010

Liu, L., Sun, L., Chen, Y., & Ma, X. (2019). Optimizing fleet size and scheduling of feeder transit services considering the influence of bike-sharing systems. *Journal of Cleaner Production*, *236*, 117550. https://doi.org/10.1016/j.jclepro.2019.07.025

Lloyd, C. D. (2010). *Local Models for Spatial Analysis*. *Local Models for Spatial Analysis*. https://doi.org/10.1201/ebk1439829196

Lu, B., Charlton, M., Harris, P., & Fotheringham, A. S. (2014). Geographically weighted regression with a non-Euclidean distance metric: A case study using hedonic house price data. *International Journal of Geographical Information Science*, *28*(4), 660–681. https://doi.org/10.1080/13658816.2013.865739

Ma, F., Shi, W., Yuen, K. F., Sun, Q., & Guo, Y. (2019). Multi-stakeholders' assessment of bike sharing service quality based on DEMATEL–VIKOR method. *International Journal of Logistics Research and Applications*, *5567*. https://doi.org/10.1080/13675567.2019.1568401

Ma, Xiaolei, Zhang, X., Li, X., Wang, X., & Zhao, X. (2019). Impacts of free-floating bikesharing system on public transit ridership. *Transportation Research Part D: Transport and Environment*, *76*(September), 100–110. https://doi.org/10.1016/j.trd.2019.09.014

Ma, Xinwei, Cao, R., & Jin, Y. (2019). Spatiotemporal clustering analysis of bicycle sharing system with data mining approach. *Information (Switzerland)*, *10*(5), 1–14. https://doi.org/10.3390/info10050163

Ma, Xinwei, Ji, Y., Yang, M., Jin, Y., & Tan, X. (2018). Understanding bikeshare mode as a feeder to metro by isolating metro-bikeshare transfers from smart card data. *Transport Policy*, *71*(November 2018), 57–69. https://doi.org/10.1016/j.tranpol.2018.07.008

McKenzie, G. (2019). Urban mobility in the sharing economy: A spatiotemporal comparison of shared mobility services. *Computers, Environment and Urban Systems*, *79*(September 2019), 101418. https://doi.org/10.1016/j.compenvurbsys.2019.101418

Metro, O. website of X. (n.d.). 首页 - 西安市轨道交通集团有限公司官网. Retrieved October 15, 2019, from https://www.xianrail.com/

Mobike. (2017). Bike-sharing and the City 2017 White Paper Introduction Bike-sharing : Making Cities More.

Modelling, W. R. (2012). GWR4 User Manual, (June 2009).

Moridpour, S., Toran Pour, A., & Saghapour, T. (2019). *Big Data Analytics in Traffic and Transportation Engineering*. IGI Global. https://doi.org/10.4018/978-1-5225-7943-4

News. (n.d.). "青桔""哈罗"相继进入 西安共享单车"投放禁令"面临尴尬. Retrieved March 2, 2020, from http://www.xinhuanet.com/tech/2018-08/22/c_1123310190.htm

O'Brien, O. (n.d.). 5.5 Million Journeys at NYC Bike Share – Suprageography. Retrieved February 7, 2020, from https://oobrien.com/2014/04/5-5-million-journeys-at-nyc-bike-share/

Oliveira, G. N., Sotomayor, J. L., Torchelsen, R. P., Silva, C. T., & Comba, J. L. D. (2016). Visual analysis of bike-sharing systems. *Computers and Graphics (Pergamon)*, *60*, 119–129. https://doi.org/10.1016/j.cag.2016.08.005

Páez, A., & Wheeler, D. C. (2009). Geographically Weighted Regression. *International Encyclopedia of Human Geography*, *28*(4), 407–414. https://doi.org/10.1016/B978-008044910-4.00447-8

Pal, A., & Zhang, Y. (2017). Free-floating bike sharing: Solving real-life large-scale static rebalancing problems. *Transportation Research Part C: Emerging Technologies*, *80*, 92–116. https://doi.org/10.1016/j.trc.2017.03.016

Qian, J., Pianura, L., & Comin, M. (2018). Data-driven smart bike-sharing system by implementing machine learning algorithms. *Proceedings - 2018 6th International Conference on Enterprise Systems, ES 2018*, 50–55. https://doi.org/10.1109/ES.2018.00015

Reiss, S., & Bogenberger, K. (2015). GPS-Data Analysis of Munich's Free-Floating Bike Sharing System and Application of an Operator-based Relocation Strategy. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, *2015-Octob*(4), 584–589. https://doi.org/10.1109/ITSC.2015.102

Reynaud, F., Faghih-Imani, A., & Eluru, N. (2018). Modelling bicycle availability in bicycle sharing systems: A case study from Montreal. *Sustainable Cities and Society*, *43*(June), 32–40. https://doi.org/10.1016/j.scs.2018.08.018

Roof, K., & Oleru, N. (2008). Public health: Seattle and king county's push for the built environment. *Journal of Environmental Health*, *71*(1), 24–27.

Shelat, S., Huisman, R., & van Oort, N. (2018). Analysing the trip and user characteristics of the combined bicycle and transit mode. *Research in Transportation Economics*, *69*(June), 68–76. https://doi.org/10.1016/j.retrec.2018.07.017

Shen, Y., Zhang, X., & Zhao, J. (2018). Understanding the usage of dockless bike sharing in Singapore. *International Journal of Sustainable Transportation*, *12*(9), 686–700. https://doi.org/10.1080/15568318.2018.1429696

Sun, Z., Li, Y., & Zuo, Y. (2019). Optimizing the Location of Virtual Stations in Free-Floating Bike-Sharing Systems with the User Demand during Morning and Evening Rush Hours. *Journal of Advanced Transportation*, *2019*. https://doi.org/10.1155/2019/4308509

Ting Ma, Chao Liu, and E. (n.d.). Bicycle Sharing and Transit: Does Capital Bikeshare Affect Metrorail Ridership in Washington, D.C.?, (301), 1–20.

Vogel, P., Greiser, T., & Mattfeld, D. C. (2011). Understanding bike-sharing systems using Data Mining: Exploring activity patterns. *Procedia - Social and Behavioral Sciences*, *20*, 514–523. https://doi.org/10.1016/j.sbspro.2011.08.058

Wang, J., Huang, J., & Dunford, M. (2019). Rethinking the utility of public bicycles: The development and challenges of station-less bike sharing in China. *Sustainability (Switzerland)*, *11*(6). https://doi.org/10.3390/su11061539

Wang, Y., Liu, Y., Ji, S., Hou, L., Han, S. S., & Yang, L. (2018). Bicycle lane condition and distance: Case study of public bicycle system in Xi'an, China. *Journal of Urban Planning and Development*, *144*(2), 1–8. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000436

Wang, Z. J., Chen, F., & Xu, T. K. (2016). Interchange between metro and other modes: Access distance and catchment area. *Journal of Urban Planning and Development*, *142*(4), 1–9. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000330

Weliwitiya, H., Rose, G., & Johnson, M. (2019). Bicycle train intermodality: Effects of demography, station characteristics and the built environment. *Journal of Transport Geography*, *74*(December 2018), 395–404. https://doi.org/10.1016/j.jtrangeo.2018.12.016

Wu, Jiansheng, Wang, L., & Li, W. (2018). Usage patterns and impact factors of public bicycle systems: Comparison between city center and suburban district in Shenzhen. *Journal of Urban Planning and Development*, *144*(3), 1–17. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000471

Wu, Jingxian, Yang, M., Sun, S., & Zhao, J. (2018). Modeling travel mode choices in connection to metro stations by mixed logit models: A case study in Nanjing, China. *Promet - Traffic - Traffico*, *30*(5), 549–561. https://doi.org/10.7307/ptt.v30i5.2623

Wu, L., Gu, W., Fan, W., & Cassidy, M. J. (2020). Optimal design of transit networks fed by shared bikes. *Transportation Research Part B: Methodological*, *131*, 63–83. https://doi.org/10.1016/j.trb.2019.11.003

Wu, X., Lu, Y., Lin, Y., & Yang, Y. (2019). Measuring the Destination Accessibility of Cycling Transfer Trips in Metro Station Areas: A Big Data Approach. *International Journal of Environmental Research and Public Health*, *16*(15), 2641. https://doi.org/10.3390/ijerph16152641

Xu, Y., Chen, D., Zhang, X., Tu, W., Chen, Y., Shen, Y., & Ratti, C. (2019). Unravel the landscape and pulses of cycling activities from a dockless bike-sharing system. *Computers, Environment and Urban Systems*, *75*(February), 184–203. https://doi.org/10.1016/j.compenvurbsys.2019.02.002

Yan, Y., Tao, Y., Xu, J., Ren, S., & Lin, H. (2018). Visual analytics of bike-sharing data based on tensor factorization. *Journal of Visualization*, *21*(3), 495–509. https://doi.org/10.1007/s12650-017-0463-1

Yang, F., Ding, F., Qu, X., & Ran, B. (2019). Estimating urban shared-bike trips with location-based social networking data. *Sustainability (Switzerland)*, *11*(11), 1–14. https://doi.org/10.3390/su11113220

Yang, Y., Heppenstall, A., Turner, A., & Comber, A. (2019). A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban flows over the last mile. *Computers, Environment and Urban Systems*, *77*(May), 101361. https://doi.org/10.1016/j.compenvurbsys.2019.101361

Yang Yongchong, LiuYing, L. L. (2018). Urban Cycling Hot Spot Evtraction Based on Sharing-bikes' Big Data, 2018.

Zhang, Lihong, Zhang, J., Duan, Z. Y., & Bryde, D. (2015). Sustainable bike-sharing systems: Characteristics and commonalities across cases in urban China. *Journal of Cleaner Production*, *97*, 124–133. https://doi.org/10.1016/j.jclepro.2014.04.006

Zhang, Lingzhu, & Zhuang, Y. (2019). Analysis of Multiple Network Accessibilities and Commercial Space Use in Metro Station Areas : An Empirical Case Study of Shanghai , China, *8*(1), 49–56.

Zhang, Y., Thomas, T., Brussel, M. J. G., & Van Maarseveen, M. F. A. M. (2017). The characteristics of bike-sharing usage: Case study in zhongshan, china. *International Journal of Transport Development and Integration*, *1*(2), 255–265. https://doi.org/10.2495/tdi-v1-n2-255-265

Zhang, Ying, Brussel, M. J. G., Thomas, T., & van Maarseveen, M. F. A. M. (2018). Mining bike-sharing travel behavior data: An investigation into trip chains and transition activities. *Computers, Environment and Urban Systems*, *69*(March 2017), 39–50. https://doi.org/10.1016/j.compenvurbsys.2017.12.004

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

Zhang, Yongping, & Mi, Z. (2018). Environmental benefits of bike sharing: A big data-based analysis. *Applied Energy*, *220*(December 2017), 296–301. https://doi.org/10.1016/j.apenergy.2018.03.101

Zhao, D., Ong, G. P., Wang, W., & Hu, X. J. (2019). Effect of built environment on shared bicycle reallocation: A case study on Nanjing, China. *Transportation Research Part A: Policy and Practice*, *128*(June), 73–88. https://doi.org/10.1016/j.tra.2019.07.018

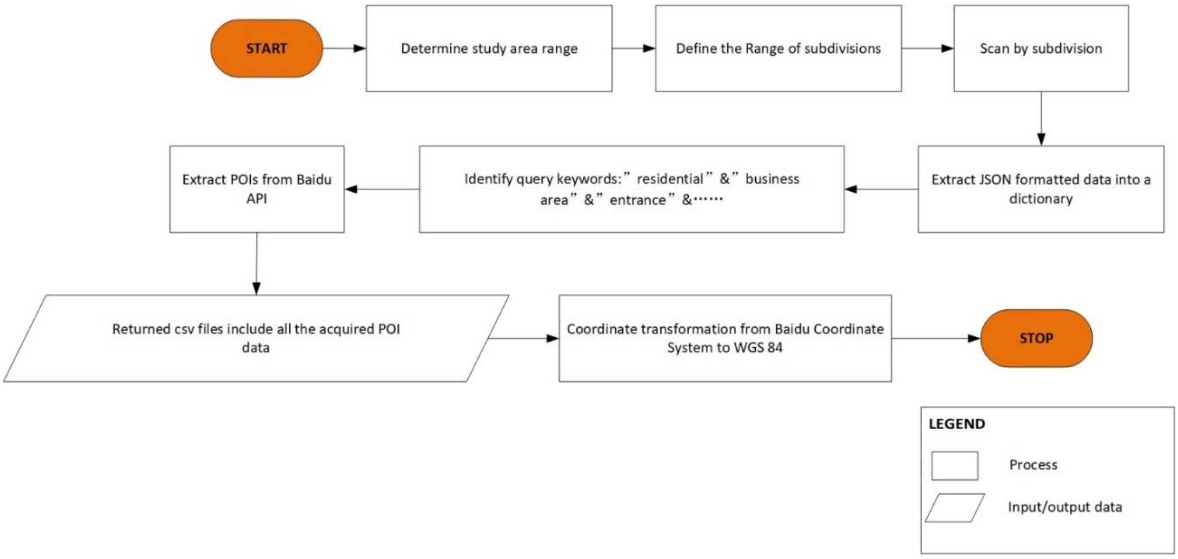
Zhao, F., Chow, L.-F., Li, M.-T., & Liu, X. (2005). A Transit Ridership Model Based on Geographically Weighted Regression and Service Quality Variables, (April), 149. https://doi.org/Final Report DO97591

Zhao, P., & Li, S. (2017). Bicycle-metro integration in a growing city: The determinants of cycling as a transfer mode in metro station areas in Beijing. *Transportation Research Part A: Policy and Practice*, *99*, 46–60. https://doi.org/10.1016/j.tra.2017.03.003

CSDN博客. (n.d.). Morans I（莫兰指数）人工智能. Retrieved April 17, 2020, from https://blog.csdn.net/allenlu2008/article/details/47130353

APPENDIX

Appendix 1: Flow chart of POI extraction



Appendix 2: The official POI classification of Baidu Map Server

Category	**Type of POI**
Job	Companies, factories and offices
residence	Residential area, dormitory
commercial	Shopping malls, department stores, supermarkets, convenience stores, building materials shop, digital shops, markets
Recreation	Resorts, cinemas, KTV, theaters, dance halls, internet cafes, game venues, bath and massage, leisure squares
Green	Parks, botanical gardens
Education	Institutions of higher learning, middle schools, elementary schools, kindergartens, adult education, parent-child education, special education schools, scientific research institutions
life services	post office, logistics company, ticket office, laundry, photo studio, real estate agency, maintenance point, housekeeping service, funeral service, lottery sales point, pet service, public toilet
Health care	General hospitals, specialty hospitals, clinics, pharmacies, medical examination institutions, emergency centers
Government	governments at all levels, administrative units, public prosecution law agencies, foreign-related agencies, party groups, welfare institutions

(Source: Baidu Map Sever)

Appendix 3: The code to determine the boundary of the dataset

bounds = [108.79, 34.13, 109.11, 34.44] (Equation1)

df=df[(df['LONGITUDE']>bounds[0])&(df['LONGITUDE']<bounds[2])& (Equation2)

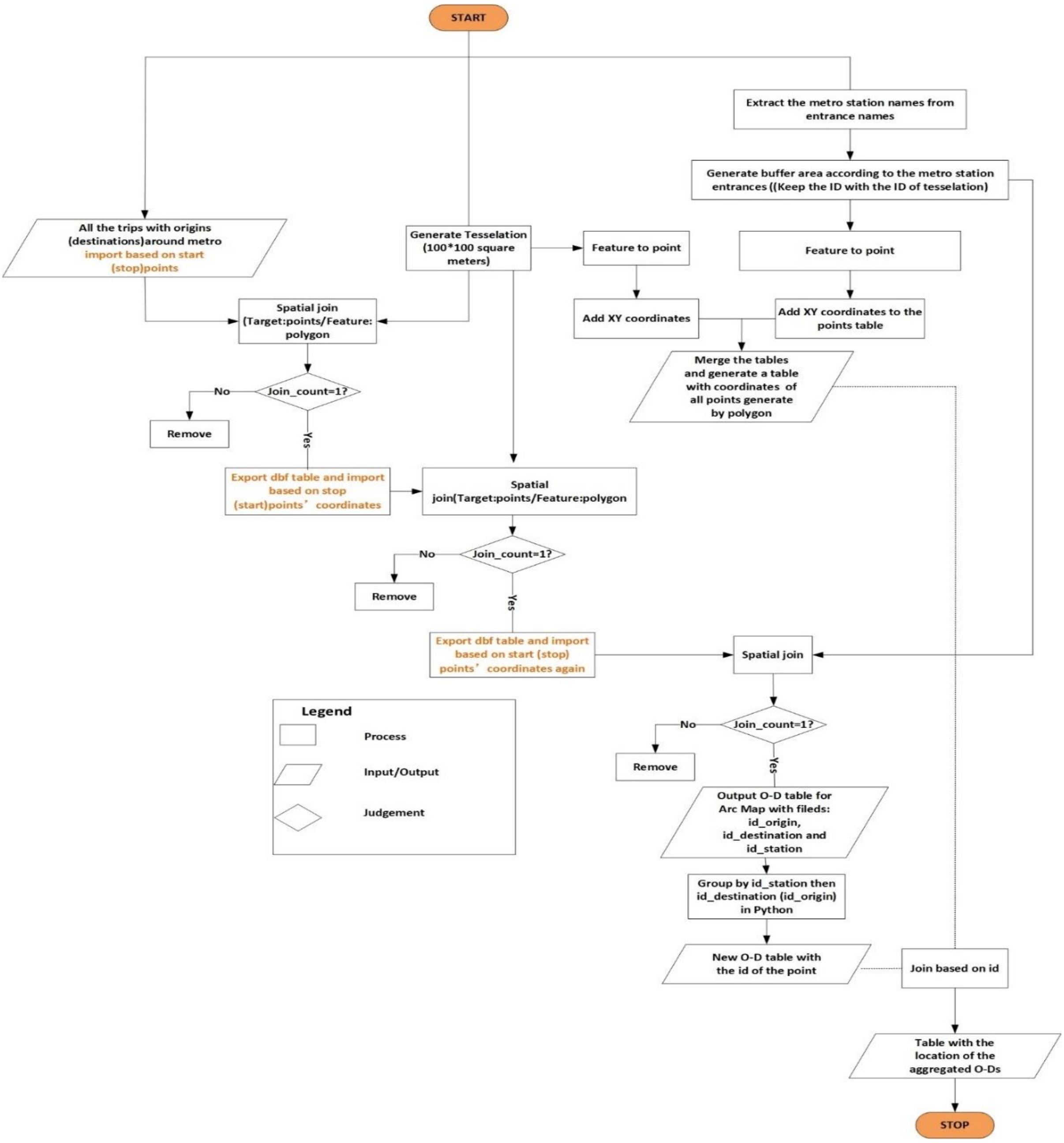
(df['LATITUDE']>bounds[1])&(df['LATITUDE']<bounds[3])]

Appendix 4: The code to calculate linear distance

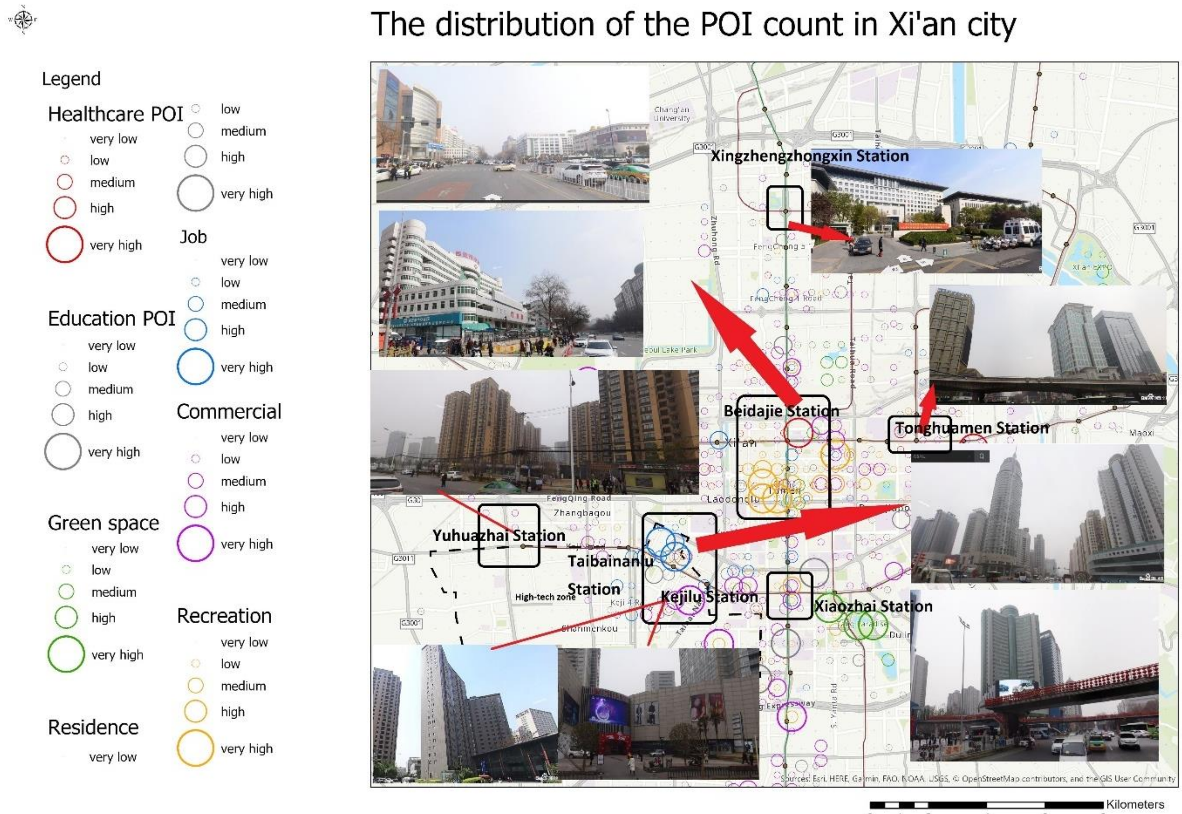
$$C = \sin(MLatA) * \sin(MLatB) * \cos(MLonA - MLonB) + \cos(MLatA) * \cos(MLatB) \quad \text{(Equation3)}$$

$$Distance = R * Arccos(C) * Pi/180 \quad \text{(Equation4)}$$

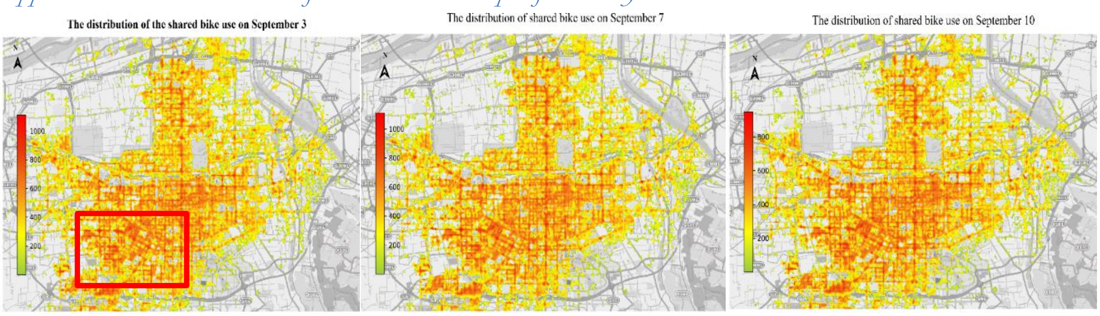
*Appendix 5: Work flow of integration the origins and destinations in 100*100 m² cells for calculating the trip count*



Appendix 6: The distribution of the POI count and the image of some hot spots.



Appendix 7: The distribution of the shared bike trips of all 3 days



Appendix 8: The iteration steps of determining the optimal bandwidth

Bandwidth search <golden section search>
Limits: 52, 1213
Golden section search begins...
Initial values

pL	Bandwidth:	52.000 Criterion:	6803.521	
p1	Bandwidth:	116.700 Criterion:	6740.757	
p2	Bandwidth:	156.687 Criterion:	6773.283	
pU	Bandwidth:	221.388 Criterion:	6831.264	
iter 1 (p1)	Bandwidth:	116.700 Criterion:	6740.757 Diff:	39.987
iter 2 (p1)	Bandwidth:	91.987 Criterion:	6727.408 Diff:	24.713
iter 3 (p2)	Bandwidth:	91.987 Criterion:	6727.408 Diff:	15.274
iter 4 (p1)	Bandwidth:	91.987 Criterion:	6727.408 Diff:	9.440
iter 5 (p1)	Bandwidth:	86.153 Criterion:	6724.889 Diff:	5.834
iter 6 (p2)	Bandwidth:	86.153 Criterion:	6724.889 Diff:	3.606
iter 7 (p1)	Bandwidth:	86.153 Criterion:	6724.889 Diff:	2.228
iter 8 (p2)	Bandwidth:	86.153 Criterion:	6724.889 Diff:	1.377
iter 9 (p1)	Bandwidth:	86.153 Criterion:	6724.889 Diff:	0.851
iter 10 (p2)	Bandwidth:	86.153 Criterion:	6724.889 Diff:	0.526

Best bandwidth size 86.000
Minimum AICc 6724.889

Appendix 9: *The geographical variability tests*

(1) In the morning peak hours, the distribution of destinations as dependent variable

Geographical variability tests of local coefficients
****

Variable	F	DOF for F test	DIFF of Criterion
Intercept	4.596016	31.992 1088.861	-76.110392
recreation	2.918736	25.984 1088.861	16.556720
commercial	1.967553	27.661 1088.861	11.192928
job	4.664692	23.805 1088.861	-59.784217
healthcare	1.697576	23.790 1088.861	16.829301
education	3.070798	25.128 1088.861	-20.195747

Note: positive value of diff-Criterion (AICc, AIC, BIC/MDL or CV) suggests no spatial variability in terms of model selection criteria.
F test: in case of no spatial variability, the F statistics follows the F distribution of DOF for F test.

(2) In the evening peak hours, the distribution of origins as dependent variable

Geographical variability tests of local coefficients
****

Variable	F	DOF for F test	DIFF of Criterion
Intercept	3.231218	35.566	-29.174673
Commercial	1.369401	28.115	35.918598
Healthcare	1.795679	26.889	21.040800
Education	0.893796	26.944	49.651965
Recreation	2.218927	28.252	8.321013
Job	3.550343	26.298	**-31.656247**

Note: positive value of diff-Criterion (AICc, AIC, BIC/MDL or CV) suggests no spatial variability in terms of model selection criteria.
F test: in case of no spatial variability, the F statistics follows the F distribution of DOF for F test.

(3) In the evening peak hours, the distribution of destinations as dependent variable

Geographical variability tests of local coefficients
****

Variable	F	DOF for F test	DIFF of Criterion
Intercept	2.836408	35.063	-13.739905
Healthcare	2.020249	26.901	14.052797
Recreation	3.839366	28.908	**-43.548350**
Residence	3.675337	28.993	**-38.497052**
Job	2.434919	25.656	1.207299

-------------------- ------------------ ---------------- -----------------

Note: positive value of diff-Criterion (AICc, AIC, BIC/MDL or CV) suggests no spatial variability in terms of model selection criteria.
F test: in case of no spatial variability, the F statistics follows the F distribution of DOF for F test.