# Location Analysis of Parcel-Pickup Points in the Guangzhou Metropolitan Area, China

April 2021

Zilai ZHENG

# Location Analysis of Parcel-Pickup Points in the Guangzhou Metropolitan Area, China

A Dissertation Submitted to

the Graduate School of Life and Environmental Sciences,

the University of Tsukuba

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in Science

(Doctoral Program in Geoenvironmental Sciences)

**Zilai　ZHENG**

# Abstract

The use of parcel-pickup points (PPPs) is an effective way of solving the last-mile problem of logistics. There are mainly two types of PPPs: parcel pick-up locker (PPL) and parcel pick-up shop (PPS). However, there are limited reports that provide decision-makers with concrete solutions for PPP organizations using quantitative methods. Where and how to select a suitable location for PPPs remains challenging. The majority of the quantitative studies focus on small residential areas, rather than metropolitan areas, as the study area. Moreover, most studies merely analyzed one type of PPP, which overlooks the interaction of the two types. This study aims to identify the location differences between the two types of PPPs in the Guangzhou metropolitan area and the layout strategy considering the interaction of the two types. The research simulated suitable location areas for the two types of PPPs with grid units, respectively. Based on the simulation results, the location differences between the two types of PPPs were analyzed. The symbiosis environment between the two types of PPPs and the same type of PPPs was analyzed in the simulated suitable location area. The study provides insights to decision-makers for rationally planning two types of PPPs to prevent mutual competition and achieve sustainability.

This study has three novel outcomes. First, ecological niche overlap theory is applied to location analysis for the sustainable development of PPPs. To avoid competition between PPPs, the two types of PPPs should be planned as a whole system when selecting sites. This study also analyzes the symbiosis environment between the two PPP types and the same PPP type. It provides reference data for locating PPPs in a suitable area. Second, this study performs a suitability simulation of PPPs in a metropolitan area using the logistic regression (LR) model of machine learning (ML). Third, this study uses detailed data to identify the specific factors for PPP locations. The study attempts to refine the population-related factors on four residential and two commercial building types; the transport infrastructure is refined to seven types of roads and three types of transportation nodes (bus stop, metro station, and parking lot).

The contributions of this study are as follows: First, the location differences between two types of PPPs were analyzed using six characteristics: the main service objects, facility attributes, impact of land price, road factors, transportation, and population. The properties of PPPs are evident based on their location characteristic. PPLs and PPSs are more inclined towards public and commercial service facilities, respectively. Second, the interaction between the two types of PPPs was clarified. PPL exhibits a high compatibility with PPS, whereas PPS exhibits a low compatibility with PPL. PPL can be considered as a supplement to PPS. Third, the LR model of the ML method performed well in both PPL and PPS. The multi-zone LR model was superior to the standard LR model. A metropolitan area is a region consisting of a densely populated urban core and its less-populated surrounding territories. Population is an important factor of PPPs. The multi-zone simulation model was preferred, and the result was more accurate in a macro-scale study area with imbalanced population. Fourth, the simulation model results revealed that the PPS area accounted for 16.5% of the total area of Guangzhou, whereas PPL accounted for 10.7%. PPP allocation focused on these suitable areas. PPP allocation significantly reduced the difficulty of the analysis and time taken during decision-making. Fifth, the structure zones impacted the PPP location. The PPP location in the three metropolitan structure zones is characterized by the fact that the most critical factor in the central zone is close to the service buildings, and the most critical factor in the middle zone and the suburban zone is close to the infrastructure.

**Keywords:** Parcel-pickup points, Parcel-pickup shop, Parcel-Pickup locker, Metropolitan structure zone, Multi-zone logistic regression model, Suitable areas, Niche overlap, Location analysis

# Contents

**Abstract** ... i

**Contents** ... iii

**List of Tables** ... vi

**List of Figures** ... vii

**Abbreviations** ... x

**Chapter 1 Introduction** ... 1

1.1 Background and problem statement ... 1

1.2 PPP distribution system ... 2

1.3 Literature review ... 8

1.4 Objectives ... 12

1.5 Research framework ... 13

1.6 Study area and data sources ... 15

1.6.1 Study area ... 15

1.6.2 Metropolitan structure zones ... 17

1.6.3 Data sources ... 23

**Chapter 2 Spatial distribution characteristic of PPPs** ... 28

2.1 Methodology ... 28

2.2 Spatial agglomeration analysis ... 31

2.2.1 Characteristics of density ... 31

2.2.2 Directional distribution ... 34

2.2.3 Agglomeration patterns ... 37

2.3 Spatial correlation analysis ... 39

2.3.1 Univariate spatial autocorrelation analysis ... 39

2.3.2 Bivariate spatial autocorrelation analysis ... 42

2.4 Discussions ... 45

**Chapter 3 Simulation of the suitable areas for PPPs** ... 47

3.1 Methodology ........................................................................................................48

3.1.1 Conversion of the multi-source data to same-scale data ........................................50

3.1.2 Creation of the PPPs database....................................................................................52

3.1.3 Diagnosis of the assumptions of LR model ..................................................................54

3.1.4 Determination of the best explanatory variable combination ......................................64

3.1.5 Evaluation of the model performance .........................................................................64

3.2 Results and discussions ...............................................................................................66

3.2.1 Optimum explanatory variable combination in standard LR model ...........................66

3.2.2 Best explanatory variable combination multi-zone LR model....................................70

3.2.3 Model performance of the standard and multi-zone LR models ................................78

3.2.4 Suitable area results by multi-zone LR model ............................................................82

3.3 Summary ......................................................................................................................86

**Chapter 4 Spatial relationship analysis of the suitable areas for PPPs** ..........................88

4.1 Methodology ................................................................................................................90

4.2 Relationship between the suitable area for the PPP types............................................93

4.3 Coexisting relationship in the overlapping area ............................................................96

4.3.1 Compatibility with intertype competition ....................................................................96

4.3.2 Compatibility with intratype competition ..................................................................100

4.4 Coexisting relationship in the non-overlap area...........................................................103

4.4.1 Distribution of the existing points in the non-overlap area on the multi-zone..........103

4.4.2 Compatibility in the non-overlap area.......................................................................107

4.5 Discussions ...............................................................................................................111

4.6 Summary ...................................................................................................................113

**Chapter 5 Layout strategy of PPPs** ..............................................................................115

5.1 Differences between PPS and PPL locations ............................................................115

5.1.1 Differences in the impact of land price .....................................................................115

5.1.2 Differences in the main service objects ....................................................................116

5.1.3 Differences in the impact of population density ........................................................116

5.1.4 Differences in the influence of road factors....117
5.1.5 Differences in the influence of transportation....118
5.1.6 Differences in facility property....118
5.2 Impact of metropolitan structure zone on PPP location....121
5.3 Layout strategy considering the mutual compatibility of the two PPP types....124
**Chapter 6 Conclusions**....127
**Acknowledgements**....132
**References**....133
**Appendices**....139
I. The explanatory variables for the suitability of PPPs....139

# List of Tables

**Page**


Table 1-1. Advantages and disadvantages of the two types of PPPs. .................................................. 7

Table 1-2. Division of the metropolitan structure zones in Guangzhou. ............................................ 20

Table 1-3. Comparison of the classes and quantities of POI data from two data sources: (a) Gaode Map data, and (b) OSM data. ........................................................................................ 25

Table 1-4. Types of roads in OSM data. ............................................................................... 26

Table 1-5. List of data used and their information. ............................................................. 27

Table 2-1. SDE information based on the types of PPPs. ................................................... 36

Table 3-1. 27 explanatory variables used in the model...................................................... 51

Table 3-2. Sample size of the observation data: (a) total data size in the positive and negative class, (b) data size of the training data and test data in PPL, and (c) data size of the training data and test data in PPS................................................................................................ 53

Table 3-3. Assumptions of the LR model. ........................................................................... 55

Table 3-4. P-values of the interaction terms in PPL and PPS. ........................................... 56

Table 3-5. VIF values of all variables after omitting the variable with the multicollinearity problem: (a) PPL data in the standard model, (b) PPS data in the standard model, (c) PPL data in the multi-zone model, and (d) PPS data in the multi-zone model. .................................................. 59

Table 3-6. Data of the outliers: (a) PPL and (b) PPS. ......................................................... 63

Table 3-7. Model performance of FSSM and BSEM in the standard LR model: (a) PPL and (b) PPS. ........................................................................................................................ 68

Table 3-8. Coefficient of the best explanatory variable combination in the standard LR model. ...... 69

Table 3-9. Coefficient of the best explanatory variable combination in multi-zone LR model for PPL. ........................................................................................................................ 76

Table 3-10. Coefficient of the best explanatory variable combination in multi-zone LR model for PPS. ........................................................................................................................ 77

# List of Figures

**Page**

Figure 1-1. Concept of distribution flow in the last-mile delivery: (a) home delivery and (b) PPP system delivery. ........ 4

Figure 1-2. Theory of the proposed study. ........ 11

Figure 1-3. Framework of the research. ........ 14

Figure 1-4. The study area, Guangzhou. ........ 16

Figure 1-5. Extension model of Guangzhou metropolitan area from the center to the periphery....... 19

Figure 1-6. Metropolitan structure zones of Guangzhou. ........ 21

Figure 1-7. Density map of transport and buildings based on the districts. ........ 22

Figure 2-1. Methodology used in Chapter 2. ........ 30

Figure 2-2. Density map based on the districts: (a) PPL, (b) PPS and (c) PPP. ........ 32

Figure 2-3. Relationship between the PPP density and population density (2010) in Guangzhou. .... 33

Figure 2-4. Directional distribution of PPPs: (a) SDE of PPL, (b) SDE of PPS, (c) SDE of PPP, and (d) overall planning of Guangzhou (2017–2035) ........ 35

Figure 2-5. KDE map: (a)PPL, and (b) PPS. ........ 38

Figure 2-6. Univariate LISA cluster distribution: (a) Distribution map of PPL, (b) Distribution map of PPS, (c)SA type % of PPL, and (d)SA type % of PPS. ........ 41

Figure 2-7. Bivariate Moran's I of PPPs with the influencing factors. ........ 44

Figure 3-1. Methodology used in Chapter 3. ........ 49

Figure 3-2. Performance gap between FSSM and BSEM in the multi-zone model (FSSM–BSEM): (a) PPL, and (b) PPS. ........ 71

Figure 3-3. Composition of the best variable combinations in the three structures: (a) PPL, and (b) PPS. ........73

Figure 3-4. Predicted performance of the final multi-zone and standard LR models in the test dataset: (a) F-measure index of PPL, (b) F-measure index of PPS, (c) Brier score of PPL, and (d) Brier score of PPS. ........ 80

Figure 3-5. Suitable area predicted by the two models in Guangzhou: (a) PPL, (b) PPS. (c) gap between two models of PPL and PPS. ........ 81

Figure 3-6. Suitability map for PPL using a multi-zone LR model: (a) central zone, (b) middle zone, (c) suburban zone, and (d) all zones. ........ 83

Figure 3-7. Suitability map for PPS using a multi-zone LR model: (a) central zone, (b) middle zone, (c) suburban zone, and (d) all zones. ........ 84

Figure 3-8. Summary of the suitable area for PPS and PPL using the multi-zone LR model. ........ 85

Figure 4-1. Different types of competitions based on the niche overlap theory. ........ 89

Figure 4-2. Methodology used in Chapter 4. ........ 91

Figure 4-3. Sample calculation of the compatibility indicators for one PPL point. ........ 92

Figure 4-4. Positional relationship between the suitable areas for PPLs and PPSs in different structure zones: (a) central zone, (b) middle zone, and (c) suburban zone, and (d) Proportion of the three types of positional relationships in the suitable areas. ........ 95

Figure 4-5. PPPs with the intertype coexisting points at the three distances of the overlapping area: (a) average number of intertype coexisting points, and (b) percentage of individuals with coexisting points. ........ 97

Figure 4-6. Box plots of the distance between the nearest points with intertype competition in the overlap area based on the different metropolitan structure zones: (a) PPS-PPL type and (b) PPL-PPS type. ........ 99

Figure 4-7. PPPs with the intratype coexisting points at the three distances of the overlap area. (a) average number of intratype coexisting points, and (b) percentage of coexisting points. ........ 101

Figure 4-8. Box plots of the distance of the nearest intratype points in the overlap area based on different metropolitan structure zones: (a) PPL-PPL and (b) PPS-PPS. ........ 102

Figure 4-9. Location and number the existing PPL points in the non-overlap area of the three structure zones: (a) central zone, (b) middle zone, and (c) suburban zone. ........ 105

Figure 4-10. Location and number the existing PPS points in the non-overlap area of the three structure zones: (a) central zone, (b) middle zone, and (c) suburban zone. .............................. 106

Figure 4-11. Box plots of the distance between the nearest intratype points in the non-overlap area based on different metropolitan structure zones: (a) PPL-PPL and (b) PPS-PPS..................... 109

Figure 4-12. PPPs with coexisting points within certain distances in the non-overlap area: (a) average number of PPPs with coexisting points, and (b) percentage of PPPs with coexisting points. .. 110

Figure 4-13. Comparing the compatibility of intertype and intratype competition in the overlapping area: (a) average number of coexisting points, and (b) average distance to the nearest points. 112

Figure 5-1. Differences between the location factors of the two PPP types in the metropolitan structure zones: (a) central zone, (b) middle zone, and (c) suburban zone. .............................. 120

Figure 5-2. Importance of the common factors of PPPs in the three metropolitan structure zones: (a) PPL and (b) PPS......................................................................................................... 122

Figure 5-3. The most important factors for PPP locations in different metropolitan structure zones. ......................................................................................................................... 123

Figure 5-4. Example of the layout strategy of PPS in the non-overlap area. .................................... 126

# Abbreviations

| | |
|---|---|
| AHP | Analytic Hierarchy Process |
| API | Application Programming Interface |
| BSEM | Backward Stepwise Elimination Method |
| CBD | Central Business District |
| DEM | Digital Elevation Model |
| FSSM | Forward Stepwise Selection Method |
| GIS | Geographic Information System |
| KDE | Kernel density estimation |
| LISA | Indicators of Spatial Association |
| LR | Logistic Regression |
| LSP | Logistics Service Providers |
| ML | Machine Learning |
| OSM | OpenStreetMap |
| POI | Point of Interest |
| PPL | Parcel-Pickup Locker |
| PPP | Parcel-Pickup Point |
| PPS | Parcel-Pickup Shop |
| SDE | Standard Deviational Ellipse |
| SA | Spatial Autocorrelation |
| VIF | Variance Inflation Factor |

# Chapter 1 Introduction

## 1.1 Background and problem statement

E-commerce has recently been rapidly increasing, resulting in changes in shopping habits. In 2017, e-commerce sales in China reached 29.16 trillion Yuan; furthermore, in 2016, the online retail e-commerce (e-retail) sales increased by 32.2% to 7.18 trillion Yuan, accounting for ~50% of the global sales. Mokhtarian (2004) reported that online shopping gives customers the ability to view product inventories and compare prices. Further, consumers are not required to physically travel to any stores and goods can be delivered directly to their homes. Compared to traditional direct-to-store delivery, the e-commerce platform provides a new, more flexible shopping option. Online shopping is not restricted to time and place. Driven by the e-commerce explosion, the volume of logistics distribution has increased precipitously. According to the National Post Bureau, China's parcel volume reached 60 billion pieces in 2019, which ranked first in the world in the volume of logistics distribution for five consecutive years from 2014. Consequently, the vigorous development of the logistics resulted in a series of environmental pollution problems such as energy consumption caused by parcel distribution, car pollutant discharge, and packaging waste.

Last-mile delivery, which is the terminal delivery in e-commerce shipping, is the most expensive, most contaminating, and least efficient component of the entire logistics process (Balcik *et al*., 2008; Cárdenas *et al*., 2017; Edwards *et al*., 2010; Gevaers *et al*., 2009; Gevaers *et al*., 2014). Last-mile delivery is considered to be the most crucial and difficult-to-control phase in e-commerce because it requires delivery of the right goods to the right place at the right time. Last-mile delivery is the final stage of the buying process, which directly affects customer satisfaction. Customer satisfaction has been on a decline owing to low service quality, such as unqualified couriers, long delivery time, and damaged parcels. Therefore, last-mile delivery hinders further development of e-commerce and logistics.

The last-mile delivery problem needs to be urgently addressed. Many e-commerce companies, logistics service providers (LSPs), and other stakeholders have considered effective delivery systems to be essential competitive advantages. Diverse innovative methods, such as parcel pick-up points (PPPs), drone delivery, and autonomous ground vehicle delivery have been proposed (Slabinac, 2015). PPPs, which are the most effective and widely used novel solution, reduce cost through consolidated shipping and provide customers with a flexible, convenient, and comfortable way of receiving parcels.

Researchers, governments, and stakeholders believe that PPPs can effectively solve the last-mile problem because all related industries cooperate to complete the PPP distribution system. This is the development trend of logistics distribution. There are several advantages of PPPs such as economic efficiency (Kämäräinen *et al*., 2001; Gevaers *et al*., 2014; Maere, 2017), environmental friendliness (Taniguchi and Kakimoto, 2003), and high service quality (Jung *et al*., 2006). The provisional regulations on express delivery issued by the Chinese government in 2018 encouraged all enterprises to share PPP facilities to provide consumers with convenient express terminal services. However, as PPPs belong to separate operating entities, they inevitably compete to gain more economic benefits. PPPs are mainly arranged based on company benefits, rather than overall planning of the entire PPP distribution system considering mutual development. Moreover, in 2020, the Coronavirus (COVID-19) pandemic changed the human lifestyle. Social distancing has limited face-to-face contact with others, resulting in more online shopping. Parcel volumes have increased rapidly. PPPs play a specific role in COVID-19 pandemic prevention and control. Therefore, how to properly arranged PPP locations is becoming particularly critical.

## 1.2 PPP distribution system

PPP distribution system can solve last-mile delivery problems, such as delivery failures, high end-to-end distribution costs, low delivery efficiency, poor service quality, and severe environmental pollution. The PPP distribution system also satisfies the diverse demands of consumers. Figure 1-1 depicts the concept of last-mile delivery based on two terminal-delivery methods— home delivery and PPP system delivery. Home delivery is the traditional method in which vehicles deliver goods from the end warehouse to each customer location. As home delivery provides door-to-door delivery services for all end customers, vehicles must be arranged to complete the delivery tasks. The entire delivery process will be affected if no one is present at the delivery location or there is a wait time for delivery. The parcels that fail first-time delivery should be returned to the warehouse for storage and added to the delivery task on the following day. The PPP delivery system is an innovative method, which adds shared PPP facilities between the terminal warehouse and the consumer. Parcels are distributed to the corresponding shared PPPs, rather than to each customer. Multiple customers that require delivery can be combined into shared PPPs. This dramatically reduces the number of delivery vehicles and the delivery workforce for enterprises, thus reducing the overall cost of last-mile logistics. In addition, traffic congestion caused by delivery vehicles occupying lanes and illegal parking is reduced, which is beneficial to city management. Environmental protection facilitates reduced exhaust emissions of delivery vehicles and improved air quality in cities (Iwan *et al*., 2016).

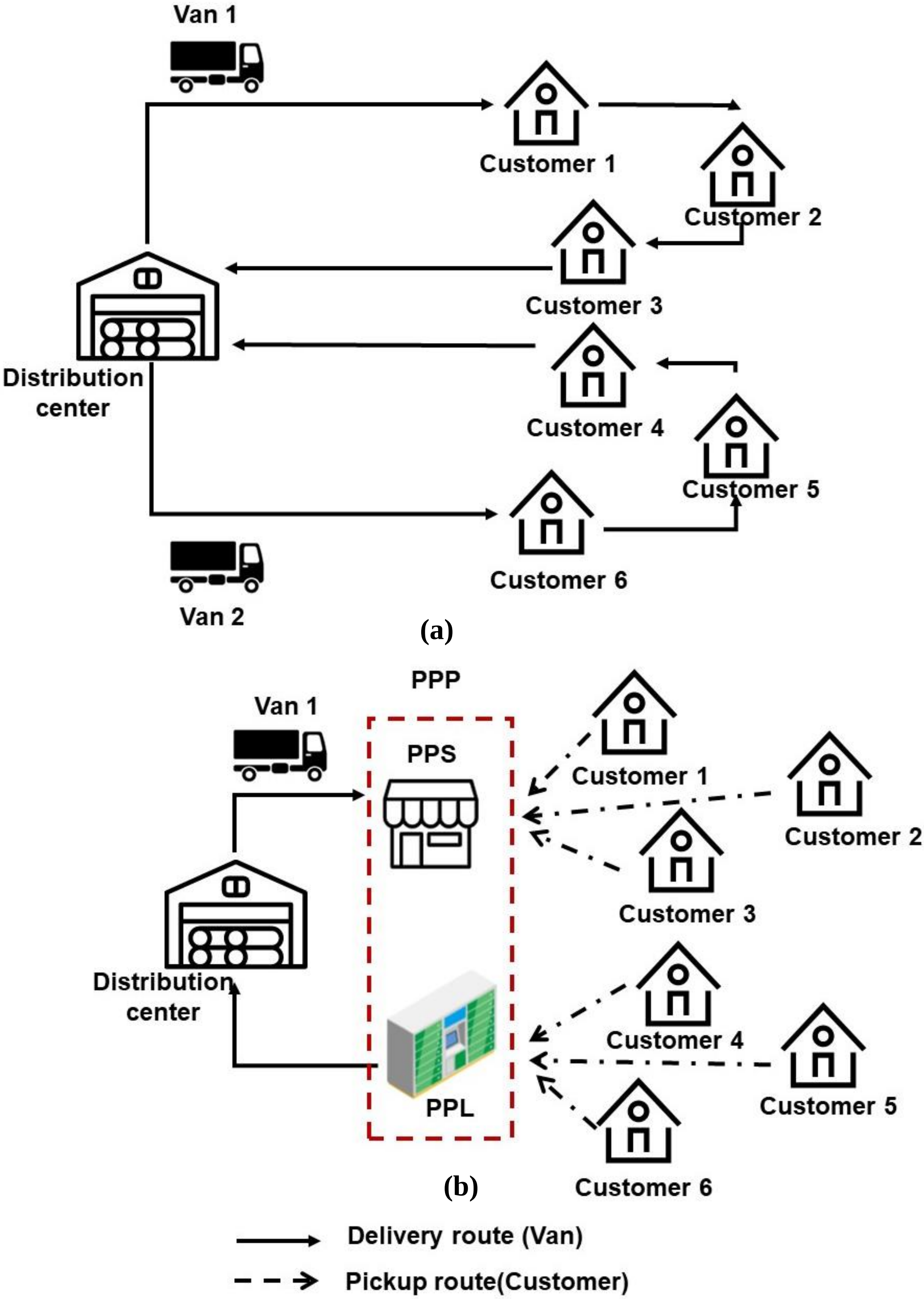



Figure 1-1. Concept of distribution flow in the last-mile delivery: (a) home delivery and (b) PPP system delivery.

There are two main types of PPPs— parcel pick-up lockers (PPLs) and parcel pick-up shops (PPSs). PPLs rely on smart technology without human interaction, whereas PPSs cooperate with commercial institutions. Both types have their advantages and disadvantages. As shown in Table 1-1, PPL exhibits advantages of opening hours, collection time, and anonymity. Consumers are allowed to collect their parcels without being bound to shop opening hours. In addition, parcels can be retrieved anonymously because no human interaction is required. However, PPLs have more disadvantages than PPSs. PPSs have more payment options, including cash payment. PPLs are not customer-friendly for specific customer groups, such as the disabled or the elderly. PPSs are more flexible owing to their ability to store parcels with different shapes and sizes without limitations of the locker size. PPLs are more prone to parcel theft because they are located in public spaces. Finally, PPSs offer face-to-face service and provide consumers with better opportunities to combine their shopping activities, because of the shop-in-shop concept. Further, they also increase the revenue of shops. For LSPs, delivering parcels to a PPS provide possibilities to combine parcel delivery with the regular store supply.

Various enterprises have established PPP distribution networks. In China, PPP systems currently form two main categories: PPL-Fengchao (Shunfeng group) and PPS-Cainiao station (Alibaba group). As PPLs and PPSs are independently planned by different enterprises, there may be duplicates in arrangement or competition with each other. This deviates from the original intention for establishing PPPs.

The attributes of PPPs have not been properly elucidated; however, they exhibit both commercial and public service attributes (Tan *et al*., 2016). The unclear attributes of PPPs cause many problems and conflicts between users and enterprises. As PPPs have not been officially defined as government public service facilities, the land for public service facilities cannot be used for PPPs. Instead, PPPs use residential communities or commercial areas by real estate developers or owners, who are willing to lease to PPP companies at a lower price and introduce PPPs to assist users because of their public service attributes. However, if the PPPs, particularly PPLs, have additional commercial activities or costs, the result will be customer dissatisfaction. At the end of April 2020, Fengchao company proposed the addition of an overtime storage fee to increase the turnover rate of the PPL facilities. Regular users can store parcels freely for 12 h. Thereafter, customers were charged 0.5 Yuan every 12 h, with a maximum of 3 Yuan. However, this became highly controversial. Public PPL facilities for residents should be nonprofit, and it is unacceptable to charge a fee. If parcels are not picked up over an extended period, the turnover rate of the enterprise will be severely affected owing to the limited use of PPL. Consequently, the PPL enterprise attempted to increase profitability to maintain the high initial investment and operating costs.

PPP delivery systems are realized by sharing receiving points, which is an important part of the popular sharing economy. The sharing economy principle increases the reuse of developed resources through the concept of sharing and reduces resource wastage and pollution. The sustainability of PPPs should also be considered to avoid competition between PPPs and the waste of resources caused by excessive allocation. Further, it should reduce the barriers between the PPP facilities as much as possible to complete a complex terminal distribution network of “PPL+PPS”.

Table 1-1. Advantages and disadvantages of the two types of PPPs.

| Item | PPL | PPS |
|---|---|---|
| **1. Opening hours** | + | - |
| **2. Collection time** | + | - |
| **3. Anonymity** | + | - |
| **4. Payment options** | - | + |
| **5. Storage possibilities** | - | + |
| **7. Security** | - | + |
| **8. Possibility to combine other shopping activities** | - | + |
| **9. Ease use** | - | + |
| **10. Face-to-face service** | - | + |

Note: "+": Strength; "-": Weakness

Source modified from Weltevreden 2008.

## 1.3 Literature review

In recent years, PPPs have attracted significant research attention in the logistics field. However, there are relatively limited empirical studies on the PPP allocation from a geographical perspective (Morganti *et al*., 2014a; Weltevreden, 2008; Zheng *et al*., 2020). Most studies are based on the extension of traditional location theory. Location theory is concerned with the geographic location of economic activity and the reason for its location.

Figure 1-2 depicts the classification of location theory. Traditional location theory is an interdisciplinary subject of economics and geography. This theory is based on the spatial distribution of human economic activities and their relationships in space. Location is the specific expression of the organic combination of physical geographic location, economic, geographic location, and geographic transportation location in space and region. The leading theory is the central place theory that seeks to explain the number, size, and location of human settlements in a residential system. The central place theory has strict assumptions: all the areas are homogeneous surfaces, and all consumers visit the nearest central places. However, the assumptions are not realistic. The extension of location theory is combined with other disciplines to supplement and improve the traditional theory. One discipline is the extension of behavioral geography. Human decision-making and the reaction to their environment are considered in the analysis. However, ignoring cost factors, location selection will damage the interests of the enterprise due to excessive emphasis on personal preferences and differences. Another discipline is the extension of quantitative geography, which quantifies various elements and relies on geographic information system (GIS) technology for analysis. However, there are limitations in the data obtained. Another discipline is structuralism, which explains the patterns of location selection. With the development of modern information technology, modern location theory relies on big data to perform cluster analysis.

Morganti *et al*. (2014b) analyzed the spatial distribution of PPPs in French urban and rural areas, and determined that population density, land use types, and transportation have an important influence on the location of PPPs. Li（2013）proposed the use of classical central place theory to identify the principle, target, and influencing factors of the PPS network layout and used a mathematical model to establish the set coverage model with the minimum number based on the PPS layout of the company. However, it did not consider factors such as market competition, regional differences, transportation accessibility, and human personal factors. From an enterprise perspective, the best location with the least transportation cost should be identified. Tan *et al*. (2016) analyzed the relationship between the spatial characteristics of the residents' behavior and PPS layout in Nanjing, China. The study revealed that the spatial layout of PPS is closely related to the social attributes, behavioral preferences, and travel modes of the residents. Lin *et al*. (2019) proposed measuring the customer's spatial accessibility to PPL considering differentiated supply and demand. Zheng *et al*. (2020) evaluated the attractiveness of different stores using resident preferences surveys and combined them with their accessibility to select the optimal candidate store as a PPS location. These studies were aimed at site selection to satisfy customers' greatest needs. Huang (2017) used a GIS platform to establish the PPS network layout model. Xue *et al*. (2019) and Li *et al*. (2018,2019) analyzed the spatial patterns and influencing factors of PPS in specific cities in China (Changsha, Xian, Wuhan) based on the point of interest (POI) data for Cainiao Stations.

The location theory in PPP allocation studies combines geography with market economic theories, regional geography, behavioral geography, quantitative geography, and clustering research. However, most of the research subjects only selected one type of PPP, and very few studies considered the relationship between PPL and PPS. In addition, the macro-scale analysis only revealed impact factors; simulations were not conducted to analyze the relationship between the factors or identify the determinate factors. Micro-scale analysis did not reveal where the candidate locations should be collected to determine the optimal location. Therefore, the simulation of the PPP suitable area can be used as a bridge between macro-scale and micro-scale analyses.

Several studies investigated the arrangement of PPP locations in specific areas, such as residential areas, campuses, and commercial areas and analyzed the characteristics of human parcel pick-up behavior in a specific environment (Lin *et al*., 2019; Tan *et al*., 2016; Zheng *et al*., 2020). The results demonstrate that the built environment affects the PPP distribution and consumer behavior. There are limited studies on macro-scale areas, such as the metropolitan area. According to the theory of urban structure, metropolitan areas can be divided into different structure zones. Population density is an important index of the metropolitan structure zones classification (Dickinson, 2013). Further, population density is also an important factor for PPP location (Morganti *et al*., 2014a). Therefore, the classification of metropolitan areas is mainly based on the population of the administrative district in this study. This study investigated whether the structure zones affect the PPP locations. And further investigations are required to individually analyze the PPP locations in specific areas inside different metropolitan structure zones in the future.

The ecology niche overlap theory is a concept of the population and community, describing how ecological objects fit together to form enduring and functioning wholes (Giller, 1984). It is widely applied in the fields of sociology and economics. The organizational niche in the sociological field considers an industry or company as a species to analyze the competitive processes and environmental dependencies. Niche marketing in the economics field considers a company, brand, or product as a species to determine its unique marketing in order to avoid competitions. Interaction analysis of PPL and PPS is attempted to base on the niche overlap theory.

Overall, there are several research gaps in the analysis of PPP locations. First, the suitability analysis of the two types of PPPs with quantitative methods was performed by limited studies, and their location differences were determined. Second, there are limited quantitative studies on large scale areas such as metropolitan areas. Third, few studies have been conducted on the overall planning of the two types of PPPs that take into account their interaction. Thus, the above-mentioned research gaps are overcome by this study, which aims to combine location theory with metropolitan structure theory and ecological overlap.

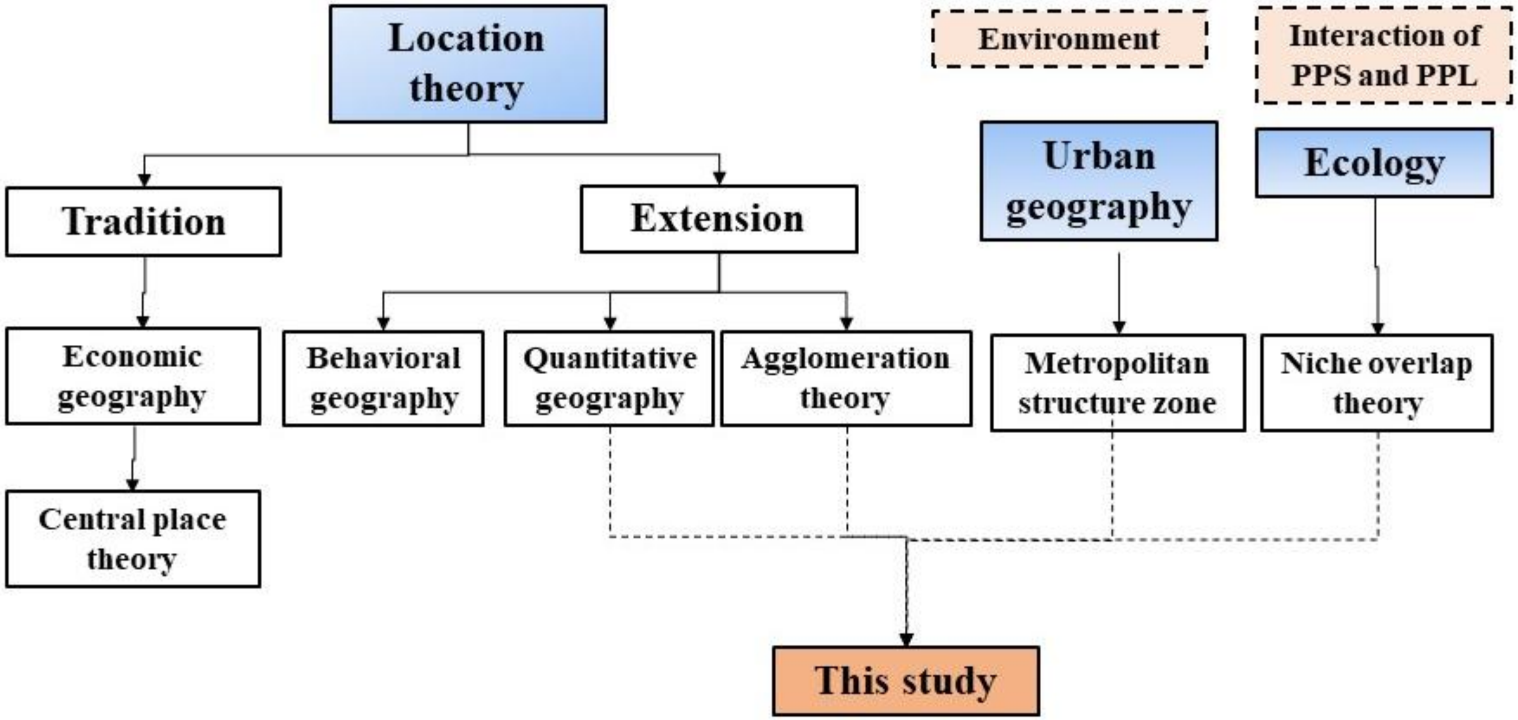



Figure 1-2. Theory of the proposed study.

## 1.4 Objectives

This research aims to identify the location differences between the two types of PPPs in the Guangzhou metropolitan area and the layout strategy considering the interaction of the two types.

In particular, the research intends to:

(1) Identify the spatial agglomeration pattern difference of the two types of PPPs under the different metropolitan structure zones.

(2) Simulate the suitable area for the two types of PPPs.

(3) Identify the determinant factors for the two types of PPPs under different metropolitan structure zones using the simulation results.

(4) Analyze the coexisting relationship between the inter- and intra-types of PPPs in the suitable area.

## 1.5 Research framework

This research is organized into six chapters to achieve its objectives, as shown in Figure 1-3. Chapter 2 describes different spatial quantitative methods to analyze the spatial agglomeration pattern and spatial correlation of PPPs. The distribution characteristics of the two types of PPPs in the three metropolitan structure zones are quite different. Chapter 3 simulates the suitable areas for the two types of PPPs based on standard and multi-zone logistic regression (LR) models using pixel units. The important factors of the two types of PPPs under the three metropolitan structure zones were determined from the 27 candidate variables using the best model. Chapter 4 analyzes the spatial relationship between the suitable areas for the two PPP types simulated in Chapter 3 and introduces the niche overlap theory to analyze their interaction. The suitability areas are divided into overlapping and non-overlapping areas. The coexisting relationship between the inter- and intra-types is explored to prevent PPP competition, and the compatibility between the two types of PPPs under the three metropolitan structure zones is discussed. In Chapter 5, the results are applied to the PPP layout strategy. The layout strategy should consider the impact of metropolitan structure zones and the mutual interaction of the two types of PPPs. Chapter 6 is conclusions.

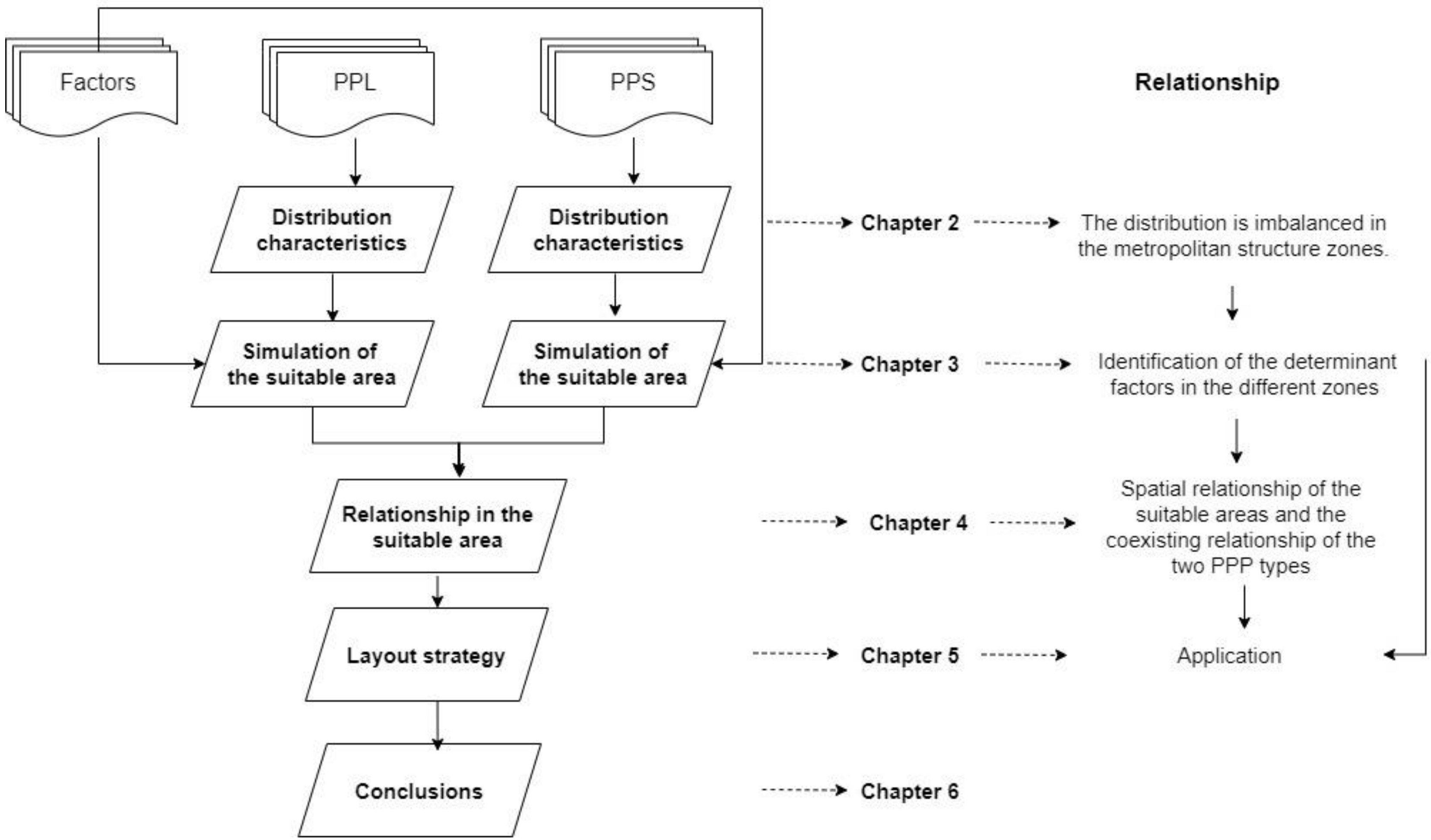



Figure 1-3. Framework of the research.

## 1.6 Study area and data sources

### 1.6.1 Study area

The study area is the Guangzhou metropolitan area, with a land area of 7434 sq. km, as shown in Figure 1-4. Guangzhou is located between 112° and 114° east longitudes, and 22° 30′ and 24° 00′ north latitudes. Guangzhou, the capital of Guangdong province with a population of 15.3 million in 2019, is adjacent to six cities: Qingyuan, Shaoguan, Foshan, Zhongshan, Dongguan, and Huizhou. Guangzhou is globally known as the southern gate of China and has the largest and oldest foreign trade port in southern China. It has an advantageous geographical location that is closest to the two special administrative regions of Hong Kong and Macau. It is the center of the Pearl River Delta Economic Zone and the Greater Bay Area, and a hub of the “One Belt, One Road”.

The choice of the study area was influenced by the following considerations:

(1) Guangzhou is one of the cities in which private LSPs occupy the market in the early stage. Many e-commerce companies in China outsource their deliveries to LSPs to reduce the cost of logistics. Given the huge parcel delivery market, numerous private LSPs were established and rapidly occupied market shares owing to their cheaper price and higher service level than the national postal services. Most of their delivery capacities were limited and covered in several coastal provinces where they initially started, mainly Zhejiang and Guangdong.

(2) Guangzhou has been ranked first in parcel receipts in China for six consecutive years from 2014 to 2019 (State Post Bureau of the People’s Republic of China,2014-2019). However, the last-mile delivery problem needs to be addressed.

(3) The number of the existing PPPs in Guangzhou is ranked second in the four most developed cities in China (Beijing, Shanghai, Guangzhou, and Shenzhen). Guangzhou was one of the first cities to establish PPP facilities.

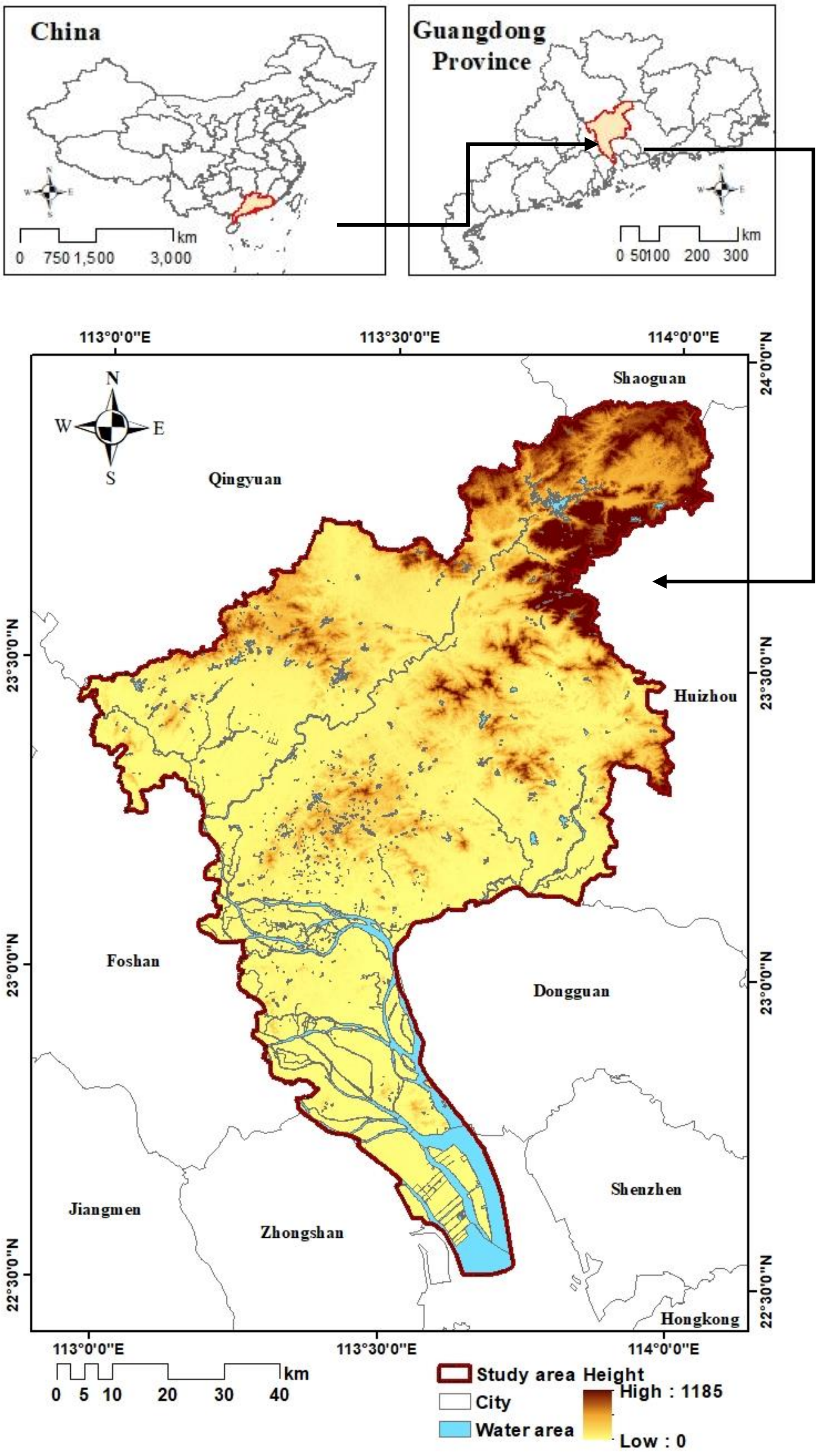



Figure 1-4. The study area, Guangzhou.

### 1.6.2 Metropolitan structure zones

A metropolitan area is a region consisting of a densely populated urban core and its less-populated surrounding territories under the same administrative division, sharing industry, infrastructure, and housing (Squires, 2002). The concentric zone theory of urban structure is the basis of the spatial structure of a city. A city grows outward from a central point in a series of concentric rings (Burgess, 1924). There are typically three zones in the model of the zonal structure— central, middle, and suburban. The structure zone has its own characteristics. The central zone is characterized by high urbanization and with the Central Business District (CBD) area. It mainly focuses on the development of the tertiary industry, and the industrial land will gradually be excluded. The middle zone is an extended area of the central zones and the land price is relatively low. The suburban zone mainly comprises leisure areas and agricultural areas on the outskirts (Dickinson, 2013).

Guangzhou metropolitan area is a typical zonal structure from the perspective of development model, population, and the degree of urbanization. The development model of the Guangzhou metropolitan area is from the center extending to the periphery, as shown in Figure 1-5. It began from Yuexiu, Dongshan (now merged into Yuexiu) district, then expanded to Haizhu and Liwan districts in the 19$^{th}$ century. In the 20$^{th}$ century, it expanded to the Tianhe and Baiyun districts at the periphery. Later, Panyu, Huangpu, Huadu, Zengcheng, and Conghua were also merged into Guangzhou metropolitan area. Now, Guangzhou becomes the third-largest metropolis of China with 11 administrative districts: Yuexiu, Liwan, Haizhu, Tianhe, Baiyun, Huangpu, Panyu, Huadu, Nansha, Zengcheng, and Conghua district. With the expansion of Guangzhou, the CBD area has also changed from Yuexiu to Tianhe district which is now the center of Guangzhou. Population density is an important index of the metropolitan area structure division (Dickinson, 2013) and also an important factor for PPP location (Morganti *et al*., 2014a). In this study, the division for a metropolitan area is mainly based on the population density in the administrative district. According to the National Bureau of Statistics of China (2010), the Guangzhou metropolitan area can be divided into three structure zones, as shown in Table 1-2 and Figure 1-6. Moreover, the urbanization degree can be reflected from the density of transport (metro station exit, bus stop, and parking lot), commercial buildings, and residential buildings, as shown in Figure 1-7. The urbanization degree of the four administrative districts in the central zone is higher than the others. The four districts in the suburban zone are with a lower density. The urbanization degree of the districts in the middle zone is between two other zones.

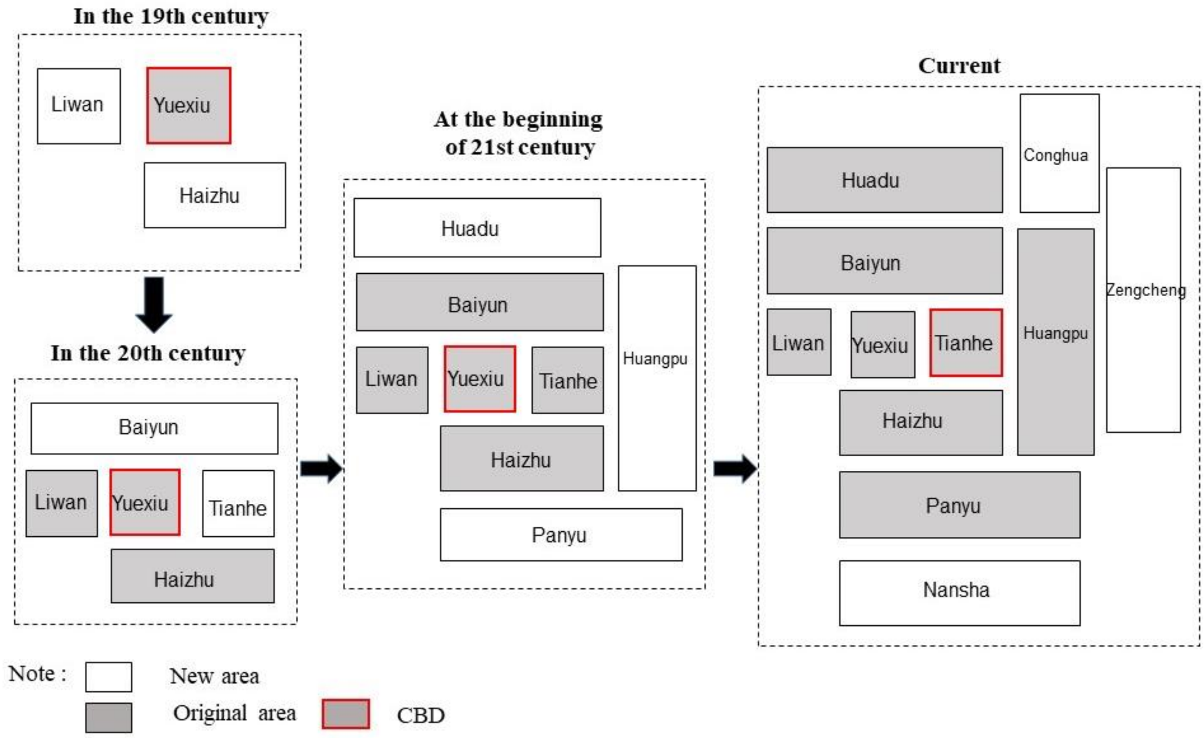



Figure 1-5. Extension model of Guangzhou metropolitan area from the center to the periphery.

Table 1-2. Division of the metropolitan structure zones in Guangzhou.

| **Structure zones** | **District** | **Area (sq. km)** | **Population density in 2010 (Persons per sq. km)** |
|---|---|---|---|
| Central zone | Yuexiu | 40.1 | 23220 |
| | Haizhu | 108 | 14439 |
| | Liwan | 74.4 | 12073 |
| | Tianhe | 161.8 | 8853 |
| Middle zone | Baiyun | 791.3 | 2810 |
| | Panyu | 615.9 | 2305 |
| | Huangpu | 574.1 | 1449 |
| Suburban zone | Huadu | 1156 | 817 |
| | Nansha | 993.6 | 609 |
| | Zengcheng | 1924.5 | 539 |
| | Conghua | 2377.1 | 250 |

Source: National Bureau of Statistics of China in 2010.

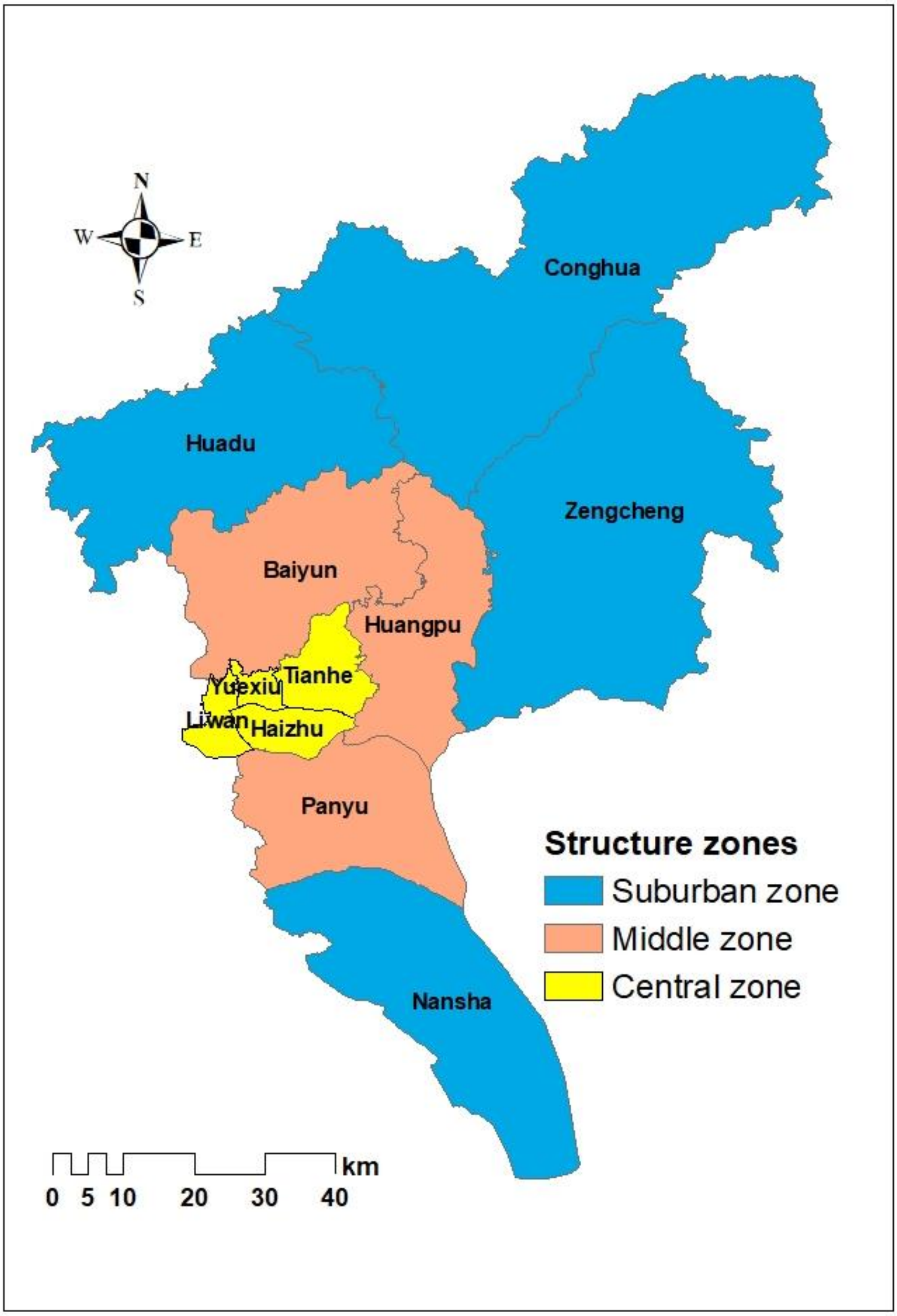



Figure 1-6. Metropolitan structure zones of Guangzhou.

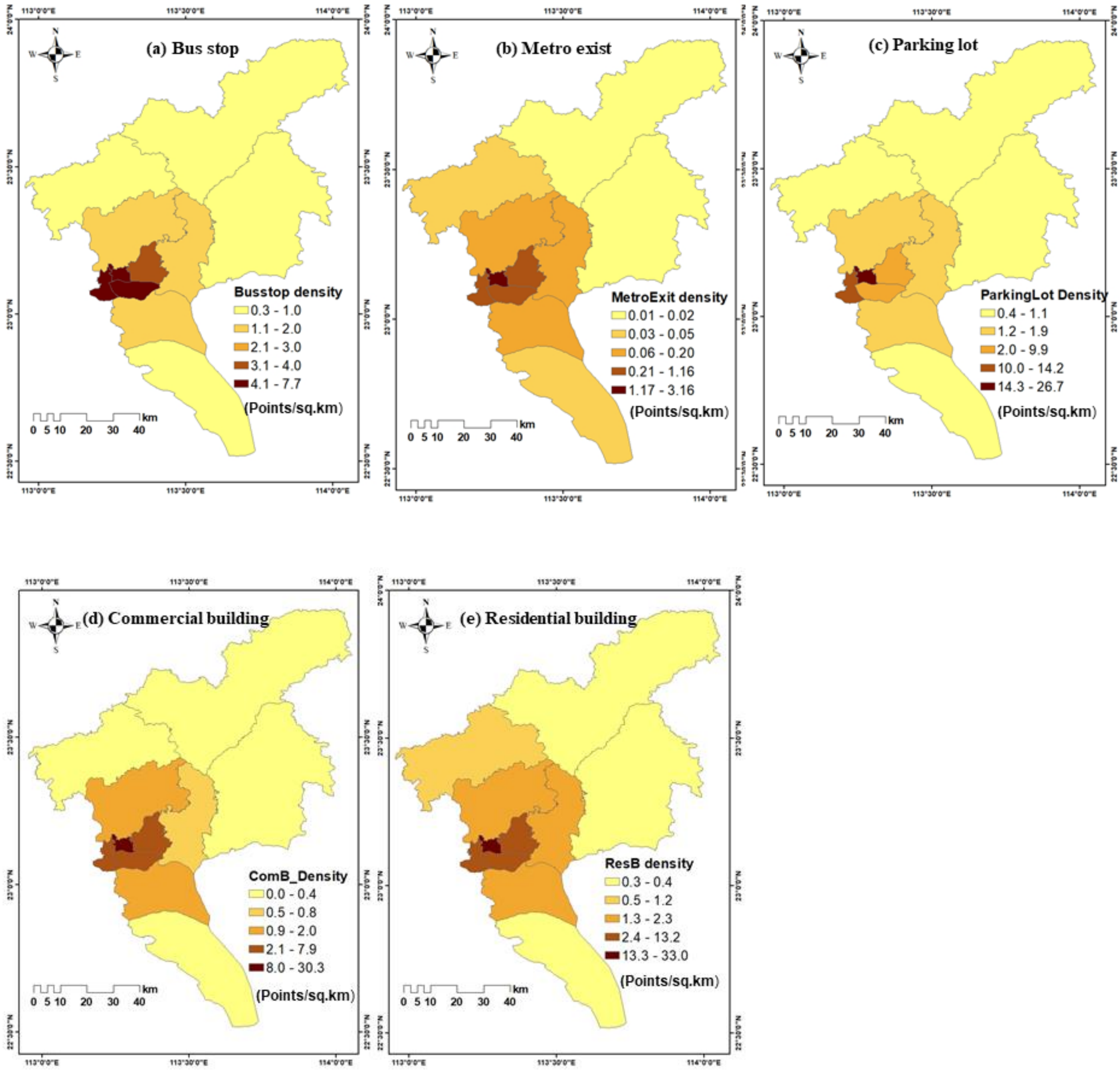



Figure 1-7. Density map of transport and buildings based on the districts.

Note: Maps created by the POI data from Gaode maps in 2019.

### 1.6.3 Data sources

This study focuses on the location analysis of two types of PPPs in Guangzhou. The primary data comprise the location of PPPs and other related facilities and roads. Point location data were collected from the POI data. POI data are novel data forms that incorporate information such as latitude and longitude coordinates, specific location, place names, and other attribute information, and play an important role in the analysis of macro-scale spatial distribution characteristics. Given the large quantity and wide distribution of PPPs, manual data acquisition is time-consuming and inaccurate. Therefore, most previous case studies have only been conducted over a small area. The difficulty involved in obtaining big data hinders the progress of PPP research. POI data have advantages of wide coverage, high recognition accuracy, and easy accessibility. The advantages of POI data are comprehensive coverage, high accuracy of identification, and easy accessibility. Thus, POI big data improve the quality of micro-scale studies on PPP locations. POI data were commonly obtained from a popular daily navigation application called Gaode Map and the crowdsourcing OpenStreetMap (OSM) project. In the Gaode Map, three-level classification codes are used to classify objects. The relevant facilities influenced by PPP mainly originated from two major categories: transportation service and commercial/house. From the open application programming interface (API) of the Gaode Map, developers can extract the specific area and classification code data. POIs provide information on latitude, longitude, location address, and shop name. If the facilities close or stop offering services, this will be revealed in the shop name. In the OSM, various types of data have been organized into point, line, and polygon data formats. POI data is one of the point data. POI information, including class and facility name, is obtained from the attribute table. According to the literature review, the distribution of PPP has a strong relationship to traffic convenience, residential and commercial areas (Morganti *et al*., 2014; Lachapelle *et al.,* 2018; Xue *et al.,* 2019). The POI data in this study will be chosen the data related to transportation, commercial and residential buildings. Table 1-3 shows the comparison of the POI data extracted from Gaode Map by the Python code and that downloaded from OSM. POI data from Gaode Map are more detailed and have higher quality than the OSM; Gaode Map was selected as the POI data source of the study. After the data cleaning, the coordinate system of the POI position in the Gaode Map needs to be converted from GCJ-02 to global WGS-84. GCJ-02 is not a coordinate system, but an algorithm that offsets the existing latitude and longitude. The conversion of the coordinate system in this study was conducted by the python code. The road data were collected from the OSM source. Table 1-4 shows the types of roads in OSM data. As the road types motorway and trunk belong to the country system and are not relevant to residents on a daily basis, these two types of roads are not considered in this study. Secondary data were gathered from the official government, including the map of the administrative area in 2017, digital elevation model (DEM) data, the standard land price for housing in 2019, and population data, as shown in Table 1-5. The population data of the administrative unit areas were obtained from the latest census conducted in 2010, and the raster data were collected from the Worldpop project in 2019.

Table 1-3. Comparison of the classes and quantities of POI data from two data sources: (a) Gaode Map data, and (b) OSM data.

(a)

| Big Category | Mid Category | Sub Category | Number |
|---|---|---|---|
| **Commercial House** | Building | Business Office Building | 5658 |
| | | Commercial-residential Building | 825 |
| | Residential Area | Villa | 280 |
| | | Residential Quarter | 7619 |
| | | Dormitory | 2031 |
| | | Community Center | 353 |
| **Transportation Service** | Subway Station | Subway Station | 317 |
| | | Exit | 808 |
| | Bus Station | Bus Station Related (Not included the airport bus stops and not operated stations) | 6778 |
| | Parking Lot | Parking Lot Related | 9882 |

(b)

| Shaple file name | Attribute table | | Number |
|---|---|---|---|
| | Fclass | Type | |
| **gis_osm_buildings** | Building | Residential | 2191 |
| | | Apartment | 1006 |
| | | Dormitory | 252 |
| | | Commercial | 733 |
| | | Office | 26 |
| **gis_osm_transport** | Bus stop | | 8291 |
| | Railway station | | 212 |

(Note: Source from OpenStreetMap website and Gaode Map API, accessed in December 2019)

Table 1-4. Types of roads in OSM data.

| Roads | Description | Note |
|---|---|---|
| **Motorway** | Include motorway, freeway, or expressway | Not considering in this study |
| **Trunk** | The most important roads in a country's system that aren't motorways. In China, it means National highway or provincial highway. | |
| **Primary** | The next most important roads in a country's system. (Often link larger towns.) | The road factor is refined to these seven road types to analyze. |
| **Secondary** | The next most important roads in a country's system. (Often link towns.) | |
| **Tertiary** | The next most important roads in a country's system. (Often link smaller towns and villages) | |
| **Unclassified** | The least important through roads in a country's system – i.e. minor roads of a lower classification than tertiary, but which serve a purpose other than access to properties. (Often link villages and hamlets.) | |
| **Residential** | Roads which serve as an access to housing, without function of connecting settlements. | |
| **Special road types** | Include living street, pedestrian, track | |
| **Paths** | Include the terms of paths and footway. | |

(Note: Source from OpenStreetMap wiki, https://wiki.openstreetmap.org/wiki/Key:highway, accessed 10.03.2021)

Table 1-5. List of data used and their information.

| Type | Layer | Description | Source | Format |
|---|---|---|---|---|
| **Spatial** | Administrative areas (boundaries) | | 1 million terrain data (2017 version) | Vector (polygon) |
| | Roads | | OSM (2019) | Vector (line) |
| | POI (Metro station exit, Bus stop, Parking lot, Residential building, Commercial building, PPS/ PPL location) | | Gaode maps (2019) | Vector(Point) |
| | DEM | DEM-GDEMV2 30 m | ASTER GDEM Project https://www.gscloud.cn/ | Raster |
| | Population | Resolution of 100 m | Worldpop Project (2019) https://www.worldpop.org/ | Raster |
| **Non-spatial** | Population census | | National Bureau of Statistics of China in 2010 | Excel |
| | Standard Land Price (Housing) | 12 price levels | Guangzhou municipal planning and natural resources bureau 2019 | Jpeg /Word |

# Chapter 2 Spatial distribution characteristic of PPPs

## 2.1 Methodology

The spatial distribution analysis of the PPPs is mainly divided into two parts—spatial agglomeration analysis of the PPPs and the spatial correlation with the factors.

The methods applied in spatial agglomeration analysis were statistical analysis, standard deviational ellipse (SDE), and kernel density estimation (KDE) methods. SDE is used to describe the directional distribution of points. The long axis orientation of the ellipse is the main distribution direction of the data set, and the short axis is the distribution range of the data set. Large differences between major and minor axes suggest more apparent data directionality (Xue *et al*., 2019). In this study, a first-order SDE containing 68% points is used as the output ellipse parameter to explore the distribution direction and the area range of PPPs in Guangzhou. KDE is an important statistical analysis method for extracting geospatial distribution characteristics and exploring the distribution pattern of points. The KDE method is based on Tobler's first law of geography.

Spatial autocorrelation (SA) can be defined as the coincidence of value similarity with locational similarity, which is used to detect patterns of spatial association (Gallo and Ertur, 2003). SA is more complex than one-dimensional autocorrelation because it is multi-dimensional and multi-directional. There are two types of SA— univariate and bivariate. Univariate SA analyzes the correlation between one variable at a location and the same variable at neighboring places. Bivariate SA analyzes the correlation between one variable at a location and another variable at neighboring places. The corresponding local indices of local Moran's I, also known as local indicators of spatial association (LISA) can display the measured value for each observation unit which enables researchers to investigate local differences in spatial dependence (Cui *et al*., 2016). Moran's I is one of the most commonly used methods for estimating SA. Moran's I is basically a Pearson correlation coefficient that uses a user-defined weight matrix and can range from – 1 to 1. Positive values for Moran's I indicate clustering, whereas negative values suggest spatial outliers. Local Moran's I is an explorative procedure and one of the most commonly used LISA that requires the global indicator Moran's I to be decomposed into each observation's contribution. Local Moran’s I disaggregates global statistics. The LISA for each observation indicates the magnitude of the substantial spatial clustering of related values around that observation. Low or high values clusters leading to positive or negative local spatial autocorrelation can be recognized on the basis of four types of spatial correlations between a given position and its neighbors (Celemín and Velázquez, 2018). Anselin (2003) created a program called GeoDa to measure global and local spatial correlation values.

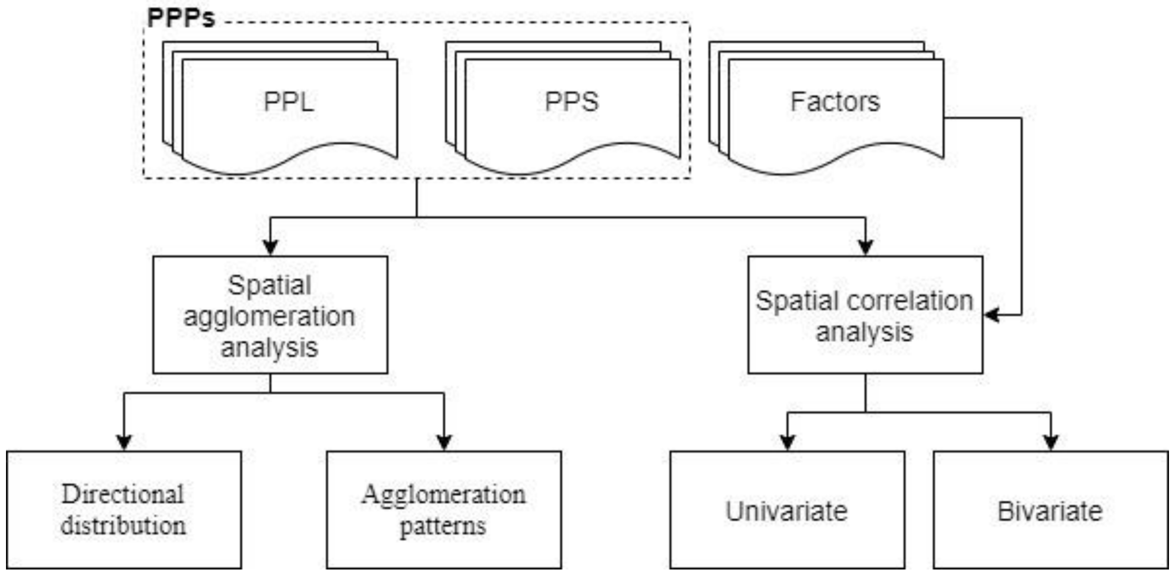



Figure 2-1. Methodology used in Chapter 2.

## 2.2 Spatial agglomeration analysis

### 2.2.1 Characteristics of density

There are 11 administration districts in the Guangzhou metropolitan area. Figure 2-2 shows the density map of PPL, PPS and PPP based on the districts, respectively. The district with the largest PPL density is Yuexiu district, which has approximately two sites per sq. km. Tianhe district had the largest PPS density, with nearly three sites in one sq. km. The top four districts with the largest PPP densities were Liwan, Yuexiu, Tianhe, and Haizhu districts, which are all in the central area of the metropolitan area. The less-dense PPP districts were Zengcheng, Huadu, Conghua, and Nansha districts, which were all distributed in the suburban area of the metropolitan area. The density values in the three districts (Baiyun, Huangpu, Panyu), located between the suburban and central areas, were also in the middle of two areas. This indicated that the value of the PPP densities decreased from the central area to the outer suburban area, which is similar to the pattern of population in the metropolitan area of Guangzhou.

Figure 2-3 shows the population density and PPP density of the districts based on their three metropolitan structure zones. The data of districts are ordering by the population density from left to right. The red points and lines represent the population density and trend of the districts. The grey, green, and blue bars represent PPP, PPL, and PPS densities, respectively. The value of PPP density basically varies by the population density of the district in Guangzhou, except for Tianhe district of the central zone. The density of PPP and population gradually decreased from the central zone, the middle area, to the suburban zone. In the three metropolitan structure zones, the characteristics of the two types of PPPs were slightly different from all PPPs. In the middle and suburban zones, the density of PPL and PPS varies with the population density of each district. The density of PPS was larger than that of PPL. In the central zone, the PPL density varied with the population density of each district. However, the PPS density was considerably different. Yuexiu had the highest population density with the lowest PPS density, whereas Tianhe had the lowest population density with the highest PPS density. There was a positive correlation between PPL density and population density. PPS density and population density were correlated in the metropolitan middle zone and the suburban zone, but not in the central zone.

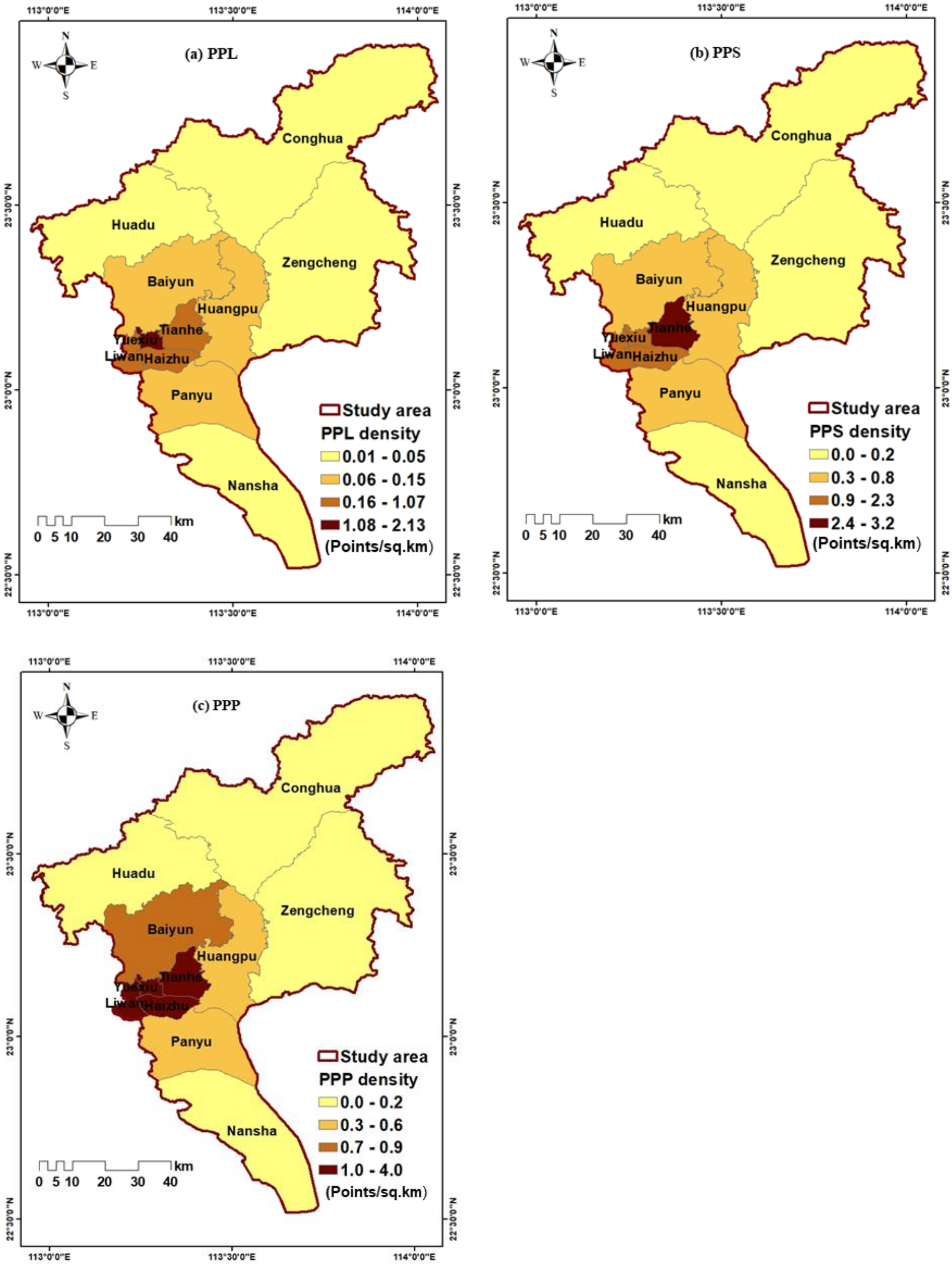



Figure 2-2. Density map based on the districts: (a) PPL, (b) PPS and (c) PPP.

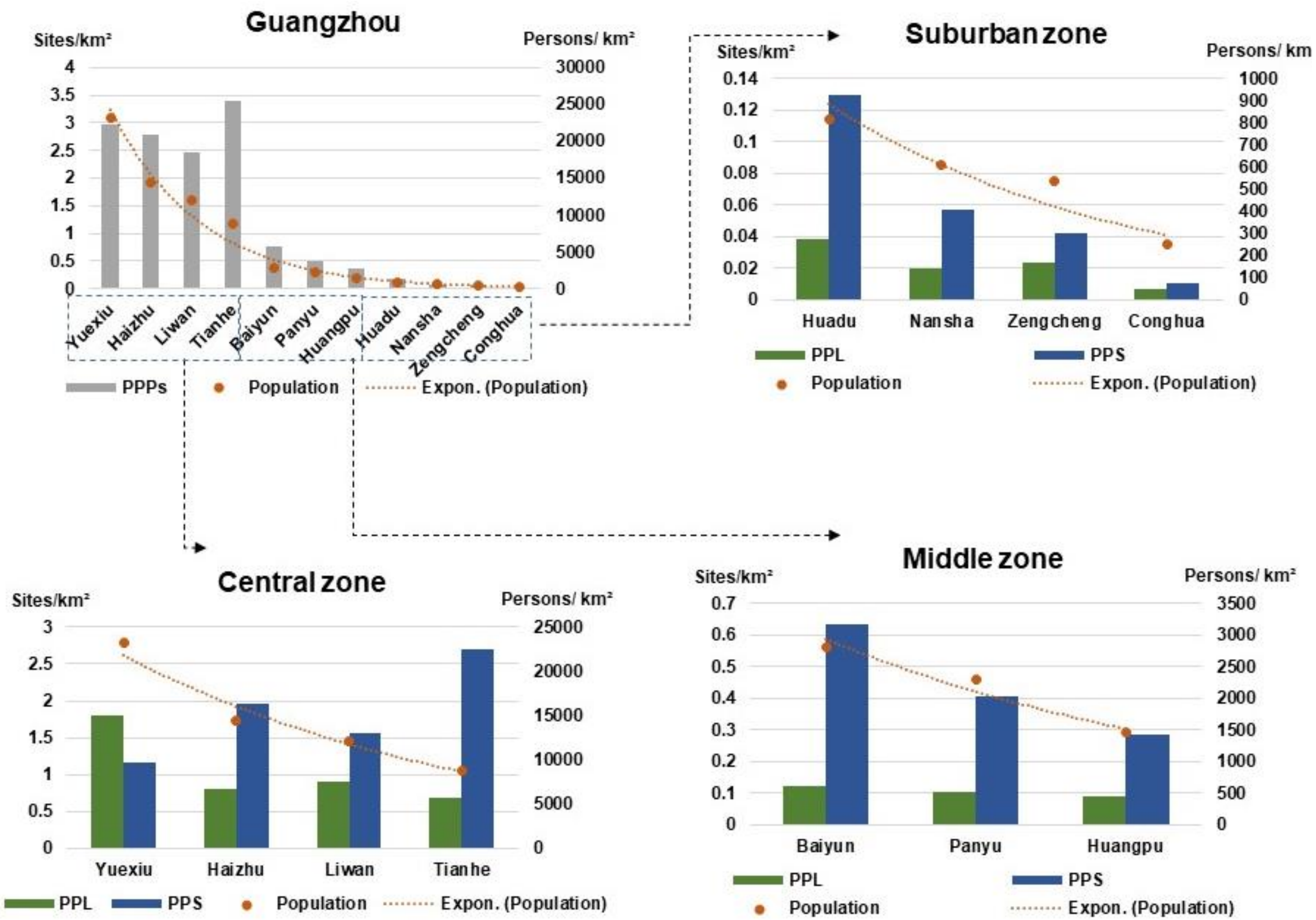



Figure 2-3. Relationship between the PPP density and population density (2010) in Guangzhou.

### 2.2.2 Directional distribution

The directional distribution of two types of PPPs is shown in Figure 2-4. The SDE of PPPs included all the areas for the central zone and part of the middle zone. Approximately 68% of PPPs were located in this area. SDE separated the middle zone into two parts. More PPPs were concentrated in the region close to the central zone, and fewer PPPs were distributed in regions further from the central zone. The middle zone exhibited mixed characteristics of the other two zones.

The PPS distribution exhibited an apparent directional trend with a 167° angular rotation (northwest-southeast). The SDE of PPL was almost a circle, without an apparent directional trend. The SDE of PPP had a 170° angular rotation, which is similar to that of PPS. The center of the three ellipses is located in the Tianhe district. Ellipses were in the seven administrative districts of Guangzhou, and the PPS area (1097 sq. km) was smaller than the PPL area (1288 sq. km).

According to the Guangzhou overall planning map from 2017 to 2023, the primary development trend of the future will extend from the central area to the Nansha sub-center in the southeast and the Huadu airport economic zone in the northwest, which is consistent with the direction of the PPP distribution. The distribution also effectively reflected the trend of the city development and can be used as a sensitive detector for monitoring metropolitan development.

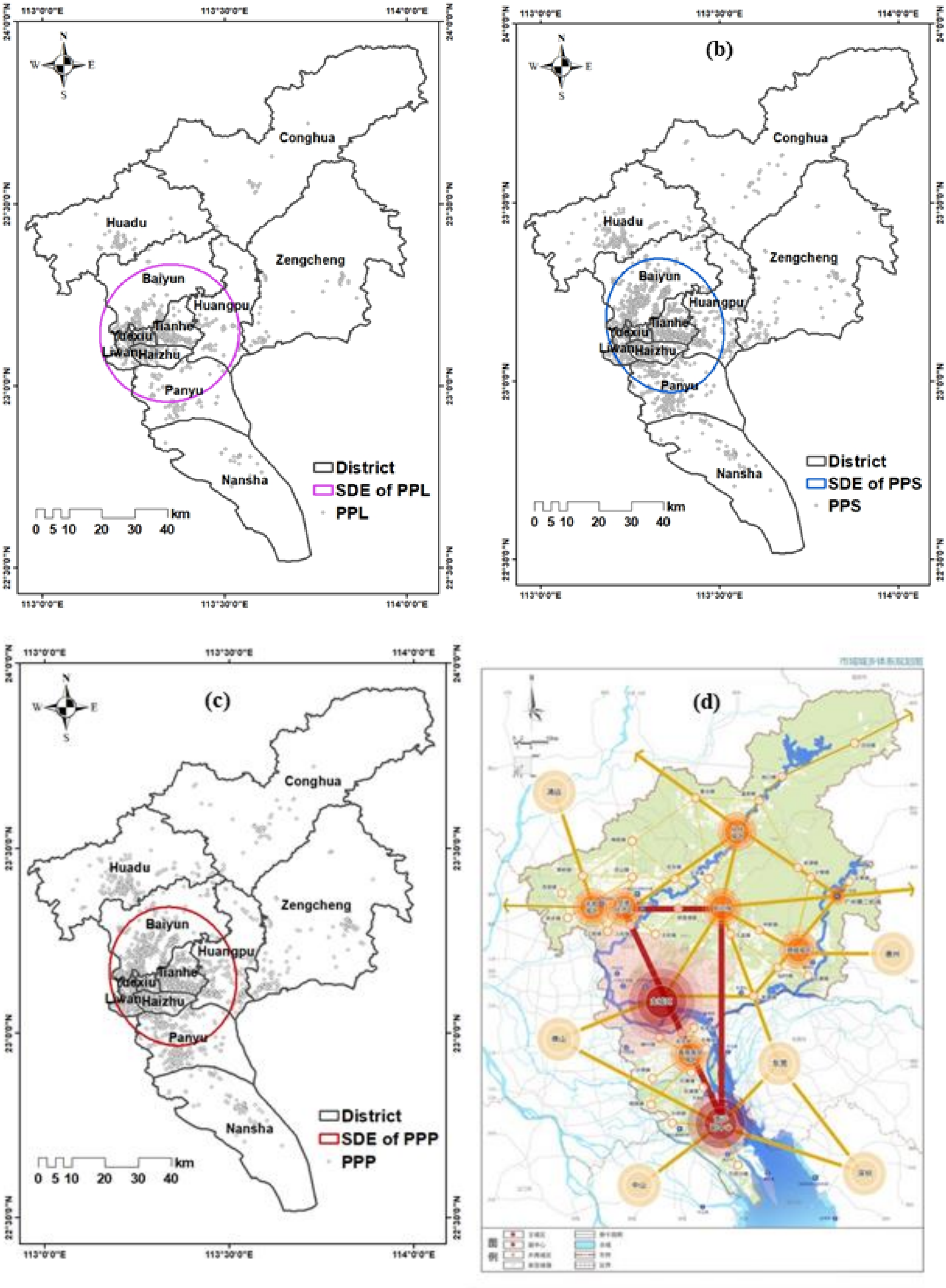



Figure 2-4. Directional distribution of PPPs: (a) SDE of PPL, (b) SDE of PPS, (c) SDE of PPP, and (d) overall planning of Guangzhou (2017–2035).

Source: Guangzhou Municipal Planning and Natural Resources Bureau, published in 2018.

Table 2-1. SDE information based on the types of PPPs.

| | **Area (sq. km)** | **Rotation (Degree)** |
| --- | --- | --- |
| SDE of PPL | 1288 | 11 |
| SDE of PPS | 1097 | 167 |
| SDE of PPP | 1150 | 170 |

### 2.2.3 Agglomeration patterns

Figure 2-5 shows the KDE of two types of PPPs. The agglomeration patterns of the two types of PPPs were significantly different. PPLs and PPSs were concentrated in single and multiple cores, respectively. The core of PPLs is mainly located in Yuexiu district, the most historic area of Guangzhou, and an extension area of the core is located in the Liwan and Haizhu districts. These three administrative districts exhibit a higher population density and longer history than the other districts of Guangzhou. PPSs are concentrated in the three cores. Two of the cores are in the central zone. One is located in the Tianhe CBD area, where many commercial and residential buildings exist. Another is located in the wholesale market area along the Xingangxi and Dongxiaonan roads of the Haizhu district. There are numerous commercial buildings and residential areas to support the wholesale market. The third core is located in the metropolitan middle zone, which is near the central zone. It is situated in the Huangbian village of the Baiyun district. There are many rental houses owing to the relatively low land price, resulting in a large population.

PPL is concentrated in areas with a long history of urban development and dense population. PPS is concentrated in CBD, wholesale commercial, and new densely populated areas. The density of the two types of PPPs in Guangzhou is inversely proportional to the distance from the core area.

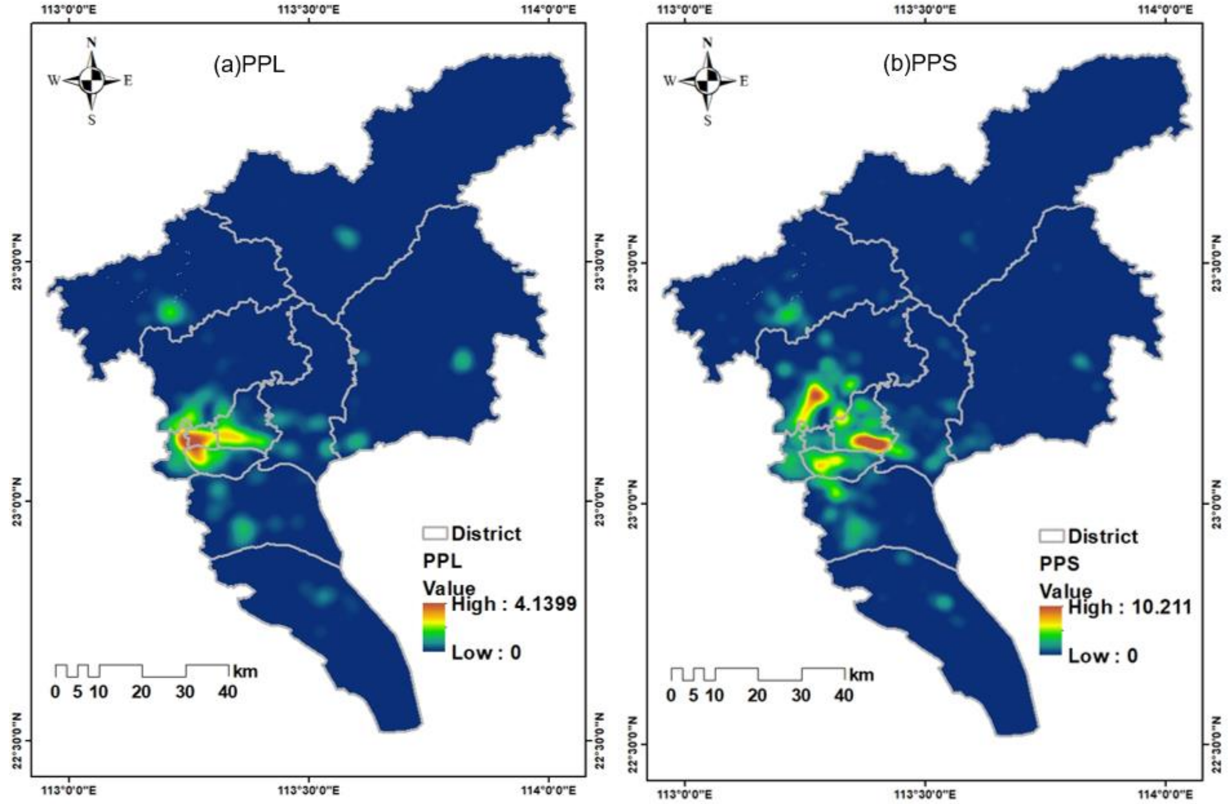



Figure 2-5. KDE map: (a)PPL, and (b) PPS.

## 2.3 Spatial correlation analysis

### 2.3.1 Univariate spatial autocorrelation analysis

Univariate SA analyzes the correlation between one variable at a location and the same variable at neighboring places. SA is more complicated than one-dimensional autocorrelation because it is multi-dimensional and multi-directional. The analysis is completed using the spatial statistics toolbox of ArcGIS 10.6 software. The parameter of the conceptualization of spatial relationships is set to the queen contiguity (contiguity_edges_corners). This parameter considers the surrounding places with shared points or common boundary conditions as the neighboring locations. When the number of administrative districts in Guangzhou is small, which cannot meet the data volume of this analysis, the smallest administrative subdistrict is used as the analysis unit. There are 167 administrative subdistricts in Guangzhou.

Moran's I is a measure of SA in statistics (Moran, 1950). From the result, the global Moran's I of PPS is 0.405, and the PPL is 0.525, which is significant under the 99.9% confidence interval. This implies that the density of two types of PPPs and that of neighboring subdistricts are positively correlated in space. The spatial distribution of two types of PPPs is clustered, and PPL has a higher degree of clustering with neighboring subdistricts than PPS.

Global Moran’s I only exhibited the overall degree of spatial aggregation and could not reflect the difference in each subdistrict. Therefore, the Anselin Local Moran’s index analysis was conducted and generated the univariate LISA cluster distribution map of the two types of density. The area was classified into five types: high-high (HH) cluster, low-low (LL) cluster, high-low (HL) outlier, low-high (LH) outlier, and not significant. The HH/LL cluster indicated that the density of PPPs in the subdistrict and neighboring subdistricts was high/low. HL/LH outlier implies that the PPP density of the subdistrict and neighboring subdistricts are high/low and low/high, respectively. Figure 2-6 shows the results of the univariate LISA cluster distribution. Most of the subdistricts belong to the LL cluster, whereas no subdistricts belong to the HL outlier group in both types of PPPs. The subdistricts with HH cluster and LH outlier are concentrated in the central zone, and the number of these subdistricts in PPS is more than that in PPL. For the PPL, there are several subdistricts with the HH type in the central zone: four in the Yuexiu district, five in the Liwan district, and three in the Haizhu district. There are only two subdistricts with LH type, namely Nonglin in

Yuexiu district and Lingnan in Liwan district. For the PPS, there are six subdistricts with the LH type in the central zone: Shijing and Tonghe in Baiyun district; Wushan and Xiancun in Tianhe district; Pazhou and Xingang in Haizhu district. There are 15 subdistricts with the HH type: two in Baiyun district, two in Liwan district, three in Haizhu district, and eight in Tianhe district. Figure 2-6 (c) and (d) show the calculated proportions of the spatial association types of subdistricts in each administrative district, and the district name was sorted based on the proportion of LL cluster type from small to large. The order of the administrative district in PPL and PPS was the same. The four administrative districts without LL cluster types were the districts in the central zone of the metropolitan area. The four administrative districts with the largest proportion of LL cluster types were in the suburban zone of the metropolitan area. The ranking of the three administrative districts in the middle zone is the same as their population density ranking. The proportion of LL type in PPL is larger than that in PPS. This indicates that the SA of PPPs is quite different in the three metropolitan structure zones. In addition, the largest proportion of HH type is in the Yuexiu district for PPL and the Tianhe district for PPS. The proportion of HH type in the Liwan district for PPL is larger than that for PPS. There is no HH type in the Baiyun district for PPL, but it does exist for PPS.

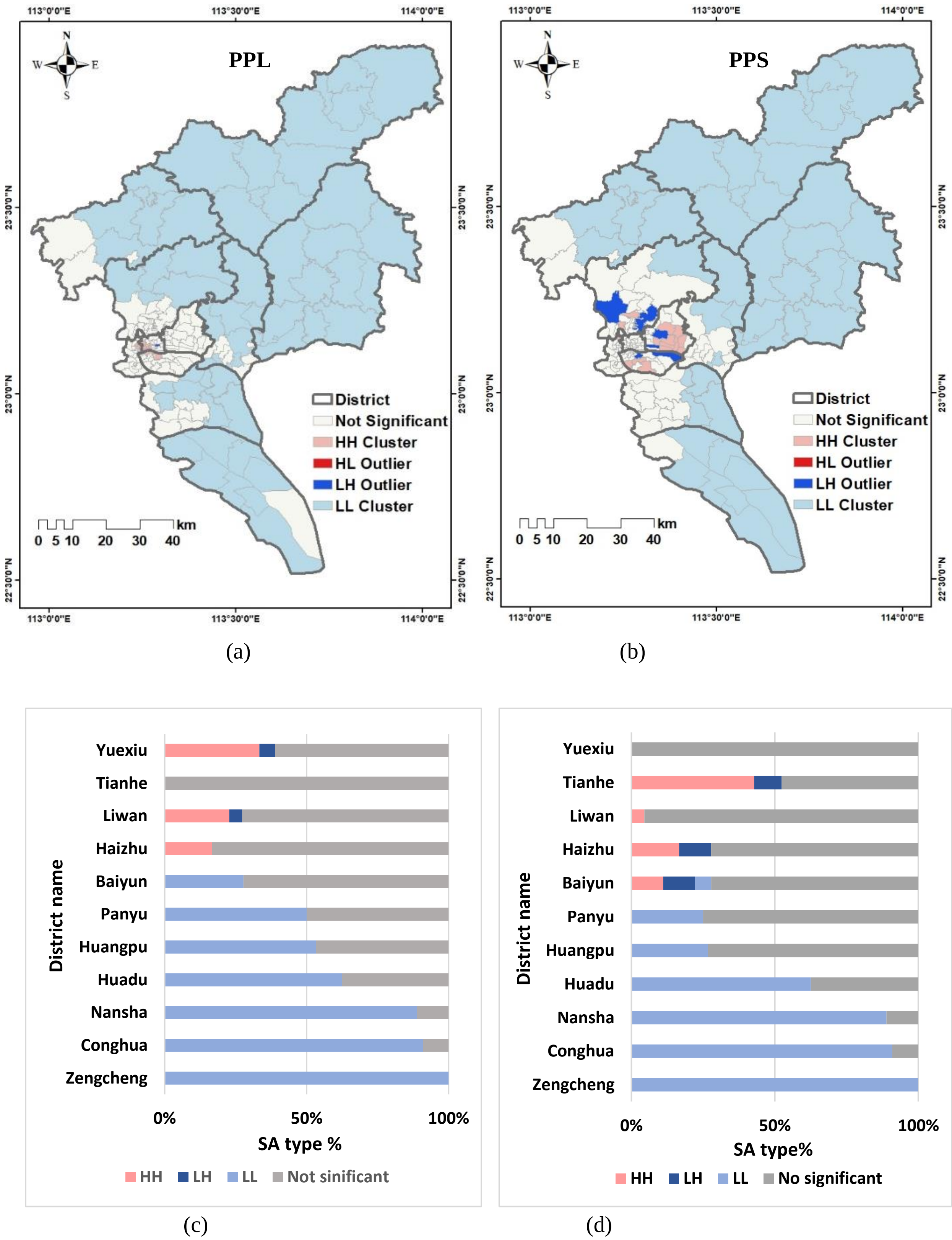



Figure 2-6. Univariate LISA cluster distribution: (a) Distribution map of PPL, (b) Distribution map of PPS, (c)SA type % of PPL, and (d)SA type % of PPS.

### 2.3.2 Bivariate spatial autocorrelation analysis

Bivariate global SA (bivariate Moran's I) explores the spatial correlation characteristics of two geographic elements, and the results represent the correlation of the overall spatial distribution of the independent variable of the region and the dependent variable of the neighborhoods. Bivariate SA was analyzed by GeoDa software in subdistrict units. The spatial relationship between the area and surrounding areas was required to be defined before analysis. As the analysis units were connected subdistricts, the contiguity weight was more suitable than the distance weight. In the contiguity weight, two criteria were selected in GeoDa software: queen contiguity and rook contiguity. The queen contiguity was more widespread, including neighbors with a common edge or vertex, whereas rook contiguity only included the shared edge. In this study, the spatial weights were constructed by queen contiguity, which is the same as the univariate SA in the previous analysis by ArcGIS software. In the 167 subdistricts of Guangzhou under the queen contiguity weight, the minimum, maximum, and mean number of neighbors were 0, 11, and 5.41, respectively.

Figure 2-7 shows the bivariate Moran's I of PPL and PPS with the seven influencing factors. The density of PPL based on the subdistrict level shows a strong positive spatial correlation with the density of the seven influencing factors of neighboring subdistricts. All the bivariate Moran's I values exceed 0.4. The seven influencing factors are arranged according to the size of the value; the order is the residential building density, population density, parking lot density, commercial building density, bus stop density, road density, and metro exit density. The largest value of the bivariate Moran's I is the residential building density (> 0.6). The smallest value is the metro exit density (= 0.4). The PPL density shows a stronger spatial relationship with the density of the residential area than with that of the commercial area, and the density of the parking lot is stronger than that of other transportation (bus stop/ metro station). For the PPS, all the values of bivariate Moran's I are less than 0.3. Therefore, the PPS density based on a subdistrict level has a weaker spatial positive correlation with the seven influencing factors of neighbors than the PPL. The seven influencing factors are arranged according to the size of the value: bus stop density, road density, population density, parking lot, commercial building density, residential building density, and metro exit density. The densities of bus stops, roads, and parking lots belonged to the factor of accessibility. Therefore, in terms of PPS, accessibility factors are more significant than the commercial and residential building factors, which is considerably different from the analysis result of PPL.

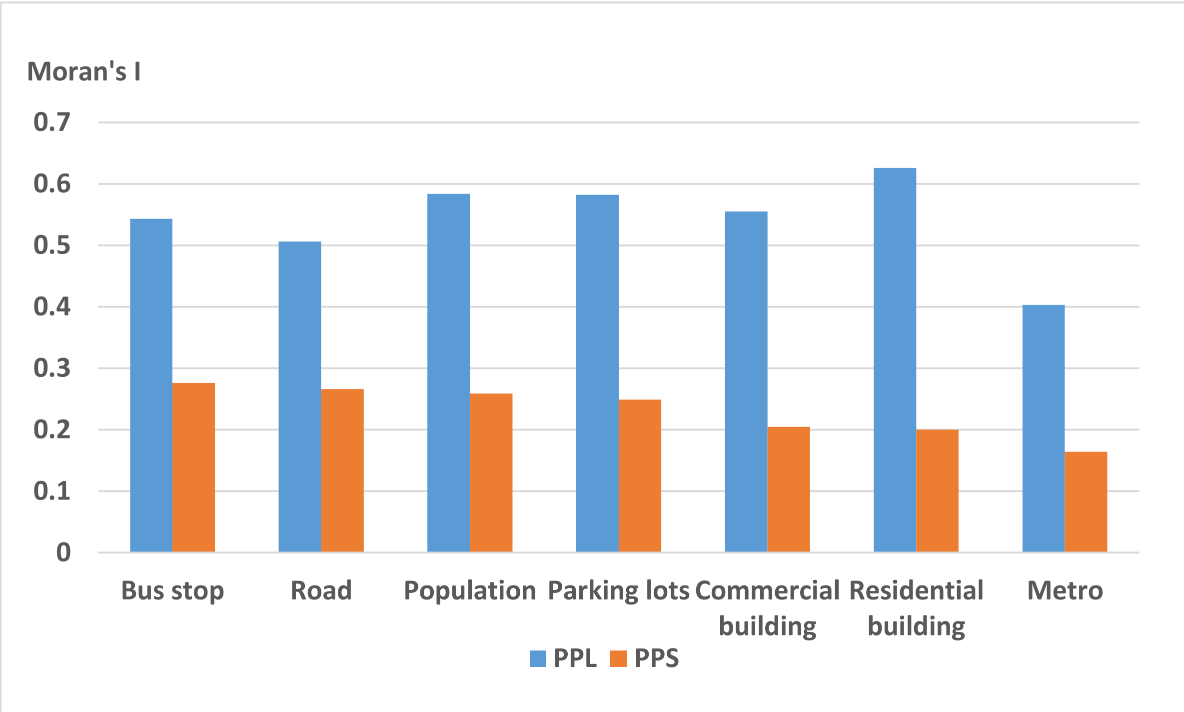



Figure 2-7. Bivariate Moran’s I of PPPs with the influencing factors.

## 2.4 Discussions

In Chapter 2, different spatial quantitative methods were applied to analyze the spatial agglomeration pattern and the spatial correlation for PPPs. The PPP distribution was imbalanced in the Guangzhou metropolitan area. The PPP distribution in the three metropolitan structural zones differed significantly. The densest area was the central zone, followed by the middle zone. A few PPPs were dispersed in the suburban zone. This is similar to the pattern of the population of Guangzhou. The distribution characteristics of PPPs reveal that density decreased from the central zone to the suburban zone. The degree of SA also gradually decreased from the suburban zone to the central zone both in PPL and PPS. In the suburban zone, the main spatial association type was the LL cluster. In the middle zone, the percentage of the LL cluster decreased. In the central zone, the HH cluster and LH outlier type were staggered, particularly the PPS. Thus, apparent spatial heterogeneity of the PPS distribution was observed in the central zone.

Moreover, the distribution characteristics of the two types of PPPs were also very different. In the administrative districts of all the structure zones, the greater the population density of the district, the greater the PPL density. However, for the PPS type, the district in the central zone did not follow that trend. Tianhe district in the central zone exhibited the largest PPS density and the lowest population density. Yuexiu district had the lowest PPS density with the highest population density. Yuexiu and Tianhe districts represent the old and new CBDs of Guangzhou, respectively. The commercial, residential, and administrative functional areas were highly concentrated in Yuexiu CBD area before. With the development of the metropolitan area, Yuexiu district outgrew its capacity. The government planned a new CBD in Tianhe where next to Yuexiu district. Many commercial functional areas in Yuexiu were transferred to the new Tianhe CBD, which retained many administrative and residential areas. Commercial areas were highly concentrated in Tianhe, while residential areas were concentrated in Yuexiu district. Tianhe CBD was one of the three national CBDs in China (the other two are Beijing CBD and Shanghai Lujiazui CBD). Given the different functions of the two districts, there would be a large difference between the daytime and nighttime populations. The population referred sourced from the government statistics bureau was a statically permanent (nighttime) population. Some researchers (Zhou *et al*., 2019) revealed that PPP density is directly related to population density and can be used to predict a population growth.

This is partly consistent with the results of this study. However, it can not be applied in the central zone for PPS, where there is a large difference between the daytime and nighttime populations. Nighttime and daytime populations may be more significant for PPL and PPS, respectively.

The agglomeration degree of PPL exceeded that of PPS. The PPS distribution had a distribution direction of northwest-southeast, whereas the PPL distribution did not exhibit a clear direction. The agglomeration pattern of PPL comprised one core mainly in the Yuexiu district of the central zone. The agglomeration pattern of PPS comprised multiple cores, two in the Tianhe and Haizhu districts of the central zone and one in the Baiyun district of the middle zone. PPL was concentrated in mature metropolitan areas with the largest population density. PPS was concentrated in CBD areas, larger wholesale commercial areas, or new densely populated areas. From the SA analysis, the PPL density exhibited a more obvious positive spatial correlation than PPS with the neighboring subdistricts. The bivariate SA analysis indicates that PPL density had a strong positive correlation with the density of the seven influencing factors of neighboring subdistricts, particularly the density of residential buildings, population, and parking lots. In comparison, PPS density exhibited a weak linear positive correlation with the seven influencing factors. The accessibility factors were more significant than the commercial or residential building factors for PPS; the opposite was true for PPL.

Overall, the PPP distribution was imbalanced in the three structure zones of the Guangzhou metropolitan area. They were mainly concentrated in the central zone, gradually decreasing from the central zone to the suburban zone. Moreover, the spatial correlation between the subdistricts in the structure zone decreased significantly from the suburban zone to the central zone, and the spatial heterogeneity of the central zone was significant. In addition, the distribution of the two types of PPPs was significantly different. The PPL density was directly proportional to the nighttime population density in districts. The PPS density was abnormal in areas with large differences in the daytime and nighttime population densities. The daytime population had a greater impact on PPS. The agglomeration degree of PPL was more significant than that of PPS. PPL density exhibited a stronger spatial positive correlation with the influencing factors of neighbors than PPS density.

# Chapter 3 Simulation of the suitable areas for PPPs

The previous chapter analyzed the PPP distribution characteristics in Guangzhou and the spatial correlation based on the subdistrict scale. The results revealed that the PPP distribution was imbalanced in the three structure zones. The two types of PPP distributions were also significantly different in the three zones. In this chapter, the distance relationship between the PPP location and impact factors will be analyzed, and the suitability will be individually simulated in the pixel scale for the two types of PPPs. Moreover, the simulation model will be separated into standard and multi-zone types and evaluate their performance.

Suitability analysis is always applied in land use or landscape to locate suitable sites for a particular purpose based on multiple criteria. Suitability analysis challenging owing to the selection of important factors and determination of their weights. For weight determination, some studies employed the Gestalt, ordinal combination, linear combination, and analytic hierarchy process (AHP) methods (Hopkins, 1977; Reed and Brown, 2003; Akıncı *et al*., 2013). The disadvantages of these methods are subjective determination of weight by humans, several restrictions of linear regression, and low efficiency. This chapter reports the use of supervised classification machine learning (ML) to address these problems. The suitability problem can be considered as a binary classification problem. The aim is to classify the suitable location pixel of PPPs. If PPP exists in the pixel, the value of the dependent variable is 1. Otherwise, it is zero. In the supervised classification algorithms, some representative sample sites are selected as training data. The characteristics of data are learned by the computer, and unknown points are classified based on the following rules: The LR model is the most common and useful model in supervised classification algorithms. It is easy to operate and efficient. The variables can be continuous and categorical. The result of the LR model is a probability from 0 to 1, which can be considered as the suitability index in this study. Therefore, the LR model was applied to simulate the suitability of two types of PPPs. For selecting factors, all the relevant factors for PPPs based on the literature review and other possibly collected environmental factors will be considered as the candidate factors in this study. The significant factors are selected based on the training data characteristics and pattern by the model.

## 3.1 Methodology

Figure 3-1 shows the methodology used in Chapter 3. It mainly comprises five steps:

(1) Conversion of multi-source data to same scale data.

(2) Preparation of the observation site data.

(3) Diagnosis of the assumptions of LR model.

(4) Determination of the best explanatory variable combination.

(5) Evaluation of the model performance.

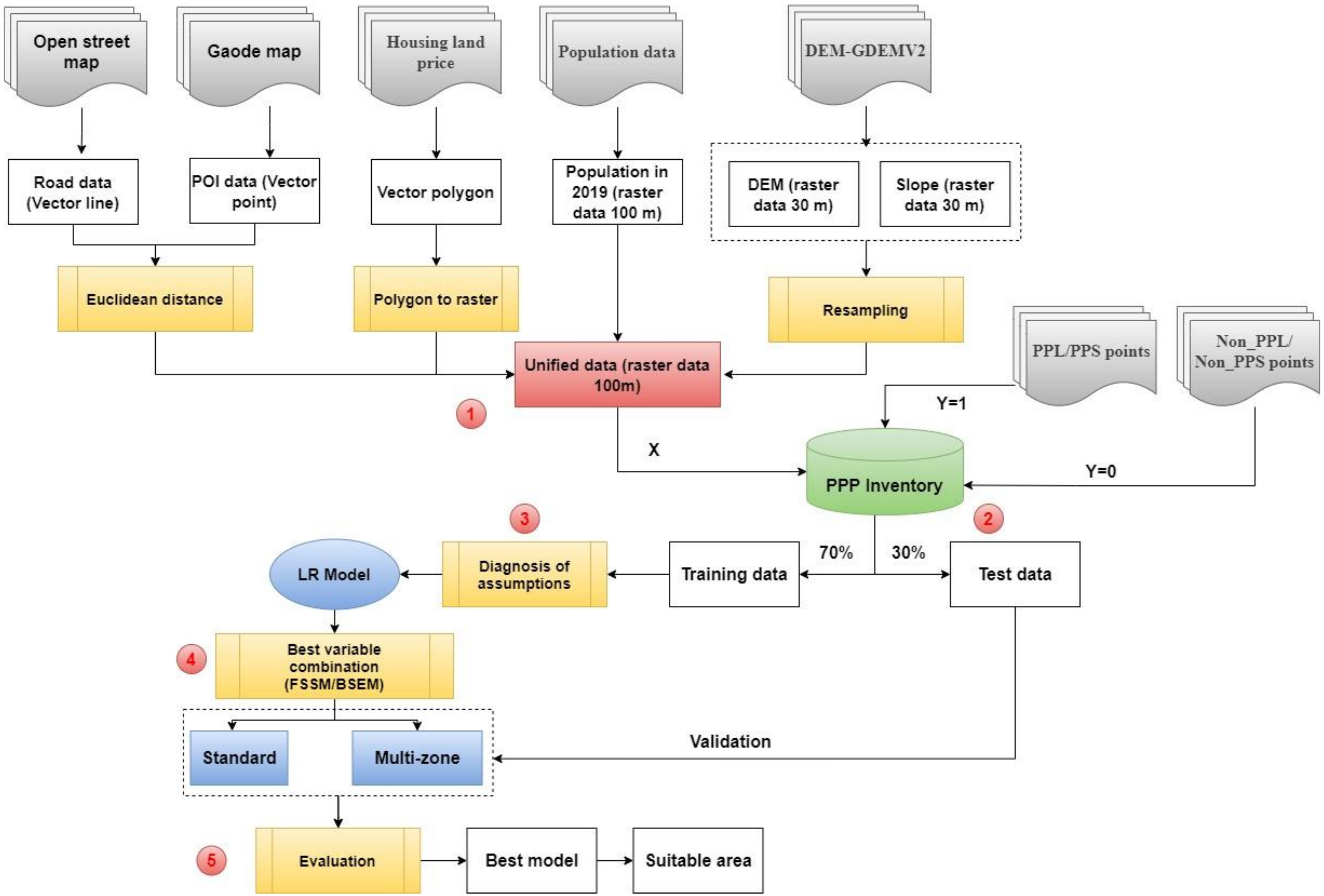



Figure 3-1. Methodology used in Chapter 3.

### 3.1.1 Conversion of the multi-source data to same-scale data

The challenges associated with multi-source data are attributed to the different types and scales of the data. All the data should be unified to the same type and unit in the pre-processing stage. There are four different data types in Chapter 3: vector line, vector point, vector polygon, and raster data with different resolutions. As this chapter aims to identify suitable areas by pixel unit, all the data need to be converted to the same raster data type and the same resolution. The vector line and point data conversion were conducted using the Euclidean distance and kernel density method. The vector polygon data were directly converted to raster data. A higher resolution of raster data was converted into a lower resolution using the resampling tool of ArcGIS10.6 software. The conversion results of all the explanatory variables are shown in Appendix I, with a resolution of 100 m. Table 3-1 shows the 27 variable names and abbreviated codes used in the model.

Table 3-1. 27 explanatory variables used in the model

| No. | Potential explanatory variable | Variable code | Unit |
|---|---|---|---|
| **1** | DEM | DEM | Number |
| **2** | Slope | Slope | Number |
| **3** | Population density | POP | Number |
| **4** | Standard land price | SLPrice | Yuan/m³ |
| **5** | Euclidean distance to the nearest residential quarter | Dist_Res_Qua | m |
| **6** | Euclidean distance to the nearest residential community center | Dist_Res_CC | m |
| **7** | Euclidean distance to the nearest residential villa | Dist_Res_Vil | m |
| **8** | Euclidean distance to the nearest residential dormitory | Dist_Res_Dor | m |
| **9** | Euclidean distance to the nearest commercial and residential building | Dist_Com_ResB | m |
| **10** | Euclidean distance to the nearest commercial office building | Dist_Com_OffB | m |
| **11** | Euclidean distance to the nearest primary road | Dist_Road_Pri | m |
| **12** | Euclidean distance to the nearest secondary road | Dist_Road_Sec | m |
| **13** | Euclidean distance to the nearest tertiary road | Dist_Road_Ter | m |
| **14** | Euclidean distance to the nearest unclassified road | Dist_Road_Unc | m |
| **15** | Euclidean distance to the nearest residential road | Dist_Road_Res | m |
| **16** | Euclidean distance to the nearest special type’s road | Dist_Road_Spe | m |
| **17** | Euclidean distance to the nearest path road | Dist_Road_Path | m |
| **18** | Euclidean distance to the nearest metro exit | Dist_MetroExit | m |
| **19** | Euclidean distance to the nearest bus stop | Dist_BusStop | m |
| **20** | Euclidean distance to the nearest parking lot | Dist_ParkingLot | m |
| **21** | Euclidean distance to the nearest water area | Dist_WaterArea | m |
| **22** | Kernel density of parking lot | Dens_ParkingLot | Number |
| **23** | Kernel density of metro exit | Dens_MetroExit | Number |
| **24** | Kernel density of bus stop | Dens_BusStop | Number |
| **25** | Kernel density of commercial building | Dens_ComB | Number |
| **26** | Kernel density of residential building | Dens_ResB | Number |
| **27** | Kernel density of road | Dens_Road | Number |

Note: Detailed data sources were shown in Table 1-5.

### 3.1.2 Creation of the PPPs database

An observation database was prepared for the LR model to learn the data features, including the suitable and unsuitable location points with their explanatory variable values. The location points of PPL and PPS were collected from the POI data of the Gaode map. This study assumes that a range of 500 m around the existing locations of PPL and PPS may still be suitable for site selection. The sampling of non-PPL and non-PPS points will be randomly generated in the remaining area after erasing the water area and the assumed suitable area. The classification ML method (particularly in the LR model) must avoid the class-imbalance problem affecting the model (Oommen *et al*.,2011). The sample size of the positive and negative data is required to be similar. As mentioned in Chapter 2, the PPP spatial location is imbalanced. Therefore, the number of unsuitable site points, when being randomly generated, was based on the number of PPL and PPS in the three metropolitan structures zones. The total number of non-PPL/PPS was almost the same as that of PPL/PPS. The details are shown in Table 3-2(a).

Next, the values of all the observation points were extracted from the 27 explanatory variable raster layers (shown in Appendices I) to create the reference database. Some points extracted the null values from the raster data of various factors: 7 points in the PPL dataset, 24 points in the non-PPL dataset, 8 points in the PPS dataset, and 99 points in the non-PPS dataset. To prevent model bias, these abnormal data were deleted. The data were randomly split into 70% training data and 30% test data according to the different PPP types and metropolitan structure zones, as shown in Table 3-2 (b) and (c).

Table 3-2. Sample size of the observation data: (a) total data size in the positive and negative class, (b) data size of the training data and test data in PPL, and (c) data size of the training data and test data in PPS.

(a)

| | **PPL** | | **PPS** | |
|---|---|---|---|---|
| | Number of PPL (1) | Number of non-PPL (0) | Number of PPS (1) | Number of non-PPS (0) |
| **Suburban zone** | 127 | 230 | 314 | 680 |
| **Middle zone** | 214 | 230 | 916 | 680 |
| **Central zone** | 338 | 230 | 811 | 680 |
| **Total** | **679** | **690** | **2041** | **2040** |

(b)

| | **PPL** | | **Non_PPL** | | **Total（PPL+Non_PPL）** | | |
|---|---|---|---|---|---|---|---|
| | Training data (70%) | Test data (30%) | Training data (70%) | Test data (30%) | Training data | Test data | ALL |
| **Suburban zone** | 88 | 38 | 158 | 68 | 246 | 105 | 351 |
| **Middle zone** | 149 | 64 | 155 | 66 | 304 | 130 | 434 |
| **Central zone** | 234 | 100 | 153 | 66 | 387 | 166 | 553 |
| **Total** | **471** | **202** | **466** | **200** | **937** | **401** | **1338** |

(c)

| | **PPS** | | **Non_PPS** | | **Total（PPS+Non_PPS）** | | |
|---|---|---|---|---|---|---|---|
| | Training data (70%) | Test data (30%) | Training data (70%) | Test data (30%) | Training data | Test data | ALL |
| **Suburban zone** | 220 | 94 | 455 | 195 | 675 | 289 | 964 |
| **Middle zone** | 638 | 273 | 447 | 192 | 1085 | 465 | 1550 |
| **Central zone** | 566 | 242 | 456 | 196 | 1022 | 438 | 1460 |
| **Total** | **1423** | **610** | **1359** | **582** | **2782** | **1192** | **3974** |

### 3.1.3 Diagnosis of the assumptions of LR model

Before applying the LR model, it is necessary to examine the seven assumptions shown in Table 3-3. Assumptions one to four are based on the dataset design and ability to satisfy the requirements, whereas assumptions five to seven require examination by other methods. Here, the diagnosis was conducted using IBM SPSS statistics 25 software.

#### *3.1.3.1 Diagnosis of the linearity of independent variables and log odds*

The Box-Tidwell method (Hosmer and Lemeshow, 1989), which incorporates the interaction term between the continuous independent variable and its natural logarithmic value into the regression equation, was employed. First, the natural logarithm of all continuous independent variables was calculated using the compute variable function in the SPSS software. Then, the interactions term between the continuous independent variable and their log was included in the binary LR analysis using SPSS. The statistical significance of this predictor suggested a non-linear logit. When the interaction term was statistically significant ($p<0.05$), there was no linear relationship between the corresponding continuous independent variable and the logit conversion value of the dependent variable. It was recommended that all items in the analysis (including the intercept term) be corrected using the Bonferroni method when testing multiple significance of the linearity hypothesis (Bland and Altman,1995). In this study, a total of 55 items were included in the model analysis, i.e., 27 continuous independent variables, 27 interaction terms with their independent variables and their natural log, and the intercept term (Constant). A significance level of 0.000091 (i.e., $0.05 \div 55$) was recommended. According to this significance level, the p-values of all interaction terms in PPL and PPS were higher than 0.000091, as shown in Table 3-4. Hence, a linear relationship exists between all continuous independent variables and the log conversion value of the dependent variable.

Table 3-3. Assumptions of the LR model.

| No | Assumptions | Explanation | Check |
|---|---|---|---|
| 1 | Dependent variable is required to be a binary variable. | Y=1: existing PPP;<br>Y=0: not exiting PPP. | N |
| 2 | Observations were required to be independent to each other. | The observations come from different measurements or matched data. | N |
| 3 | There is at least one dependent variable. The independent variable can be a continuous variable or a categorical variable. | One dependent variable and 27 continuous independent variables. | N |
| 4 | Large size of the sample is required. A general guideline is that a minimum sample quantity of 10 times the number of your model's independent variables is needed. | 27 independent variables. At least 270 cases of data. The number of PPS points dataset is 14025 and PPL dataset 1205. | N |
| 5 | The linearity of independent variables and log odds is assumed. | Box-Tidwell method | Y |
| 6 | There is little or no multicollinearity among the independent variables. | Multicollinearity diagnosis | Y |
| 7 | There are no obvious outliers' points. | | Y |

Table 3-4. P-values of the interaction terms in PPL and PPS.

| No. | Interaction terms | P-value (PPL) | P-value (PPS) |
|---|---|---|---|
| **1** | Dist_BusStop by ln_Dist_BusStop | 0.999 | 0.8781 |
| **2** | Dist_Com_OffB by ln_Dist_Com_OffB | 0.998 | 0.7625 |
| **3** | Dist_Com_ResB by ln_Dist_Com_ResB | 0.998 | 0.2346 |
| **4** | Dist_MetroExit by ln_Dist_MetroExit | 0.997 | 0.922 |
| **5** | Dist_Road_Path by ln_Dist_Road_Path | 0.998 | 0.4369 |
| **6** | Dist_Road_Pri by ln_Dist_Road_Pri | 0.995 | 0.0003 |
| **7** | Dist_Road_Res by ln_Dist_Road_Res | 0.997 | 0.0934 |
| **8** | Dist_Road_Secd by ln_Dist_Road_Sec | 0.999 | 0.5097 |
| **9** | Dist_Road_Spe by ln_Dist_Road_Spe | 0.998 | 0.1153 |
| **10** | Dist_Road_Ter by ln_Dist_Road_Ter | 1 | 0.2659 |
| **11** | Dist_Road_Unc by ln_Dist_Road_Unc | 1 | 0.3829 |
| **12** | Dist_ParkingLot by ln_Dist_ParkingLot | 1 | 0.6148 |
| **13** | Dist_Res_CC by ln_Dist_Res_CC | 0.999 | 0.1442 |
| **14** | Dist_Res_Dor by ln_Dist_Res_Dor | 1 | 0.3423 |
| **15** | Dist_Res_Qua by ln_Dist_Res_Qua | 0.995 | 0.7181 |
| **16** | Dist_Res_Vil by ln_Dist_Res_Vil | 0.998 | 0.1312 |
| **17** | POP by ln_POP | 0.998 | 0.9433 |
| **18** | Slope by ln_slope | 1 | 0.0437 |
| **19** | SLPrice by ln_ SLPrice | 0.997 | 0.8864 |
| **20** | Dist_WaterArea by ln_ Dist_WaterArea | 1 | 0.3526 |
| **21** | Dens_ResB by ln_ Dens_ResB | 0.999 | 0.0042 |
| **22** | Dens_Road by ln_ Dens_Road | 1 | 0.1932 |
| **23** | Dens_ComB by ln_ Dens_ComB | 0.999 | 0.2662 |
| **24** | Dens_BusStop by ln_ Dens_BusStop | 0.999 | 0.0111 |
| **25** | Dens_MetroExit by ln_ Dens_MetroExit | 0.997 | 0.947 |
| **26** | Dens_ParkingLot by ln_ Dens_ParkingLot | 1 | 0.0373 |
| **27** | DEM by ln_DEM | 0.999 | 0.9552 |

*3.1.3.2. Diagnosis of multicollinearity*

The best LR model exhibits low noise and is statistically robust. Therefore, the explanatory variables were highly correlated with the dependent variable, but minimally correlated with each other (Midi *et al*., 2010). Multicollinearity occurred when the explanatory variables exhibited a strong correlation or association with each other. When the degree of correlation was extremely high, the standard errors of the coefficients increased, which caused some variables to appear statistically insignificant in the result, even though they were significant. Multicollinearity made the coefficients unstable (Belsley *et al*., 1980), and reduced the precision or interfered with the result when fitting the model (Schroeder *et al.*,1990). It was mainly detected with the help of tolerance (Tol) and reciprocal, called the variance inflation factor (VIF) (Mansfield and Helms,1982). The formulae are defined as follows:

$$\mathrm{Tol} = 1 - R^2 \tag{1}$$

$$\mathrm{VIF} = 1/\mathrm{Tol} = 1/\,(1 - R^2) \tag{2}$$

where $R^2$ is the coefficient of determination for the regression of the explanatory variable on all remaining independent variables.

VIF > 10 and tolerance < 0.1 are a common threshold for assessing multicollinearity between the explanatory variables (Kroll and Song, 2013; Midi *et al*., 2010). There are several ways to address the multicollinearity problem. First, multiple variables that are collinear can be combined into a single variable. Second, the sample size can be increased to decrease standard errors. Third, some variables causing multicollinearity may be omitted from the model. Omitting some variables is the most direct, simple, and effective way. The function of collinearity diagnostics in SPSS generated the result with two tables: collinearity statistics and variance proportions table. To maintain as many variables as possible and determine their' importance, the most correlated variable was deleted each time until the collinearity problem was not severe.

Table 3-5 shows the VIF value of all variables after omitting the variable with multicollinearity in the standard and multi-zone models. In the standard model, 2 variables were deleted, and 25 variables remained for both the PPL and PPS data. The PPL variables of Dens_ResB and Dens_MetroExit were deleted, and the PPS variables of Dist_MetroExit and Dist_Com_ResB were deleted. For the PPL data in the multi-zone model, three variables were deleted from the central zone: Dens_ParkingLot, Dens_MetroExit, and

Dens_ResB; one variable was removed from the middle zone: density of Dens_ResB; two variables were omitted from the suburban zone: Dens_BusStop and Dens_ParkingLot. For the PPS data in the multi-zone model, two variables were deleted from the central zone: Dens_ParkingLot and Dens_MetroExit; one variable was removed from the middle zone: density of Dens_ResB; three variables were omitted from the suburban zone: Dist_MetroExit, Dist_Com_ResB, and Dens_BusStop.

*3.2.3.3. Outlier detection*

An outlier is an exceptional value that is far different from the others in a dataset. The LR model is sensitive to outliers. The usual approach for detecting outliers is determined using the value of standardized residuals. An absolute value larger than three is usually considered an outlier (Hosmer and Lemeshow, 1989). Table 3-6 shows 5 and 15 outliers in the PPL and PPS training datasets, respectively. After deleting the outliers, model fitting was conducted in the training dataset of 961 PPL samples and 2968 PPS samples.

.

Table 3-5. VIF values of all variables after omitting the variable with the multicollinearity problem: (a) PPL data in the standard model, (b) PPS data in the standard model, (c) PPL data in the multi-zone model, and (d) PPS data in the multi-zone model.

(a)

| | | **Step0** | | **Step1** | | **Step2** | |
|---|---|---|---|---|---|---|---|
| **No.** | Variable | Tol | VIF | Tol | VIF | Tol | VIF |
| **1** | Dist_WaterArea | 0.693 | 1.444 | 0.695 | 1.438 | 0.698 | 1.433 |
| **2** | POP | 0.682 | 1.466 | 0.692 | 1.445 | 0.692 | 1.444 |
| **3** | Slope | 0.594 | 1.685 | 0.594 | 1.683 | 0.595 | 1.681 |
| **4** | Dist_Road_Unc | 0.559 | 1.788 | 0.560 | 1.786 | 0.560 | 1.785 |
| **5** | Dist_Res_Vil | 0.533 | 1.878 | 0.533 | 1.877 | 0.557 | 1.797 |
| **6** | Dist_Road_Pri | 0.452 | 2.212 | 0.458 | 2.182 | 0.458 | 2.181 |
| **7** | Dist_Road_Res | 0.424 | 2.359 | 0.425 | 2.356 | 0.425 | 2.353 |
| **8** | Dist_BusStop | 0.380 | 2.630 | 0.381 | 2.623 | 0.382 | 2.618 |
| **9** | Dist_Road_Ter | 0.375 | 2.664 | 0.377 | 2.650 | 0.378 | 2.643 |
| **10** | Dist_Road_Spe | 0.326 | 3.066 | 0.327 | 3.062 | 0.328 | 3.052 |
| **11** | Dist_Road_Path | 0.320 | 3.123 | 0.324 | 3.091 | 0.325 | 3.079 |
| **12** | DEM | 0.316 | 3.164 | 0.316 | 3.163 | 0.316 | 3.161 |
| **13** | Dist_Road_Sec | 0.305 | 3.282 | 0.307 | 3.256 | 0.309 | 3.241 |
| **14** | Dist_Res_CC | 0.292 | 3.419 | 0.293 | 3.419 | 0.296 | 3.376 |
| **15** | Dist_Res_Dor | 0.183 | 5.475 | 0.183 | 5.458 | 0.183 | 5.454 |
| **16** | SLPrice | 0.165 | 6.044 | 0.166 | 6.022 | 0.203 | 4.918 |
| **17** | Dens_ComB | 0.160 | 6.251 | 0.178 | 5.632 | 0.211 | 4.741 |
| **18** | Dist_MetroExit | 0.155 | 6.452 | 0.155 | 6.452 | 0.159 | 6.271 |
| **19** | Dist_Com_ResB | 0.151 | 6.637 | 0.151 | 6.626 | 0.154 | 6.490 |
| **20** | Dist_Res_Qua | 0.142 | 7.021 | 0.143 | 6.984 | 0.143 | 6.969 |
| **21** | Dist_Com_OffB | 0.140 | 7.142 | 0.140 | 7.130 | 0.140 | 7.128 |
| **22** | Dens_ParkingLot | 0.139 | 7.196 | 0.173 | 5.769 | 0.175 | 5.704 |
| **23** | Dist_ParkingLot | 0.110 | 9.098 | 0.110 | 9.096 | 0.110 | 9.095 |
| **24** | Dens_Road | 0.106 | 9.401 | 0.110 | 9.087 | 0.119 | 8.395 |
| **25** | Dens_BusStop | 0.098 | 10.235 | 0.108 | 9.241 | 0.114 | 8.756 |
| **26** | Dens_MetroExit | 0.097 | 10.286 | 0.097 | 10.283 | Deleted | |
| **27** | Dens_ResB | 0.086 | 11.685 | Deleted | | | |

(b)

| | | **Step0** | | **Step1** | | **Step2** | |
|---|---|---|---|---|---|---|---|
| **No.** | **Variable** | **Tol** | **VIF** | **Tol** | **VIF** | **Tol** | **VIF** |
| **1** | Slope | 0.698 | 1.433 | 0.698 | 1.433 | 0.695 | 1.438 |
| **2** | Dist_Road_Unc | 0.574 | 1.743 | 0.574 | 1.742 | 0.579 | 1.727 |
| **3** | POP | 0.552 | 1.811 | 0.552 | 1.811 | 0.548 | 1.826 |
| **4** | Dist_WaterArea | 0.531 | 1.882 | 0.533 | 1.875 | 0.529 | 1.892 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **5** | Dist_Res_Vil | 0.47 | 2.128 | 0.483 | 2.069 | 0.489 | 2.043 |
| **6** | Dist_Road_Ter | 0.44 | 2.275 | 0.44 | 2.272 | 0.441 | 2.268 |
| **7** | Dist_ BusStop | 0.423 | 2.367 | 0.436 | 2.292 | 0.432 | 2.316 |
| **8** | Dist_ Road_Pri | 0.356 | 2.807 | 0.361 | 2.773 | 0.361 | 2.771 |
| **9** | Dist_ Dist_Road_Res | 0.307 | 3.255 | 0.315 | 3.177 | 0.317 | 3.157 |
| **10** | Dist_ Road_Spe | 0.283 | 3.529 | 0.286 | 3.502 | 0.293 | 3.408 |
| **11** | DEM | 0.271 | 3.692 | 0.271 | 3.688 | 0.286 | 3.491 |
| **12** | Dist_ Road_Path | 0.261 | 3.824 | 0.262 | 3.817 | 0.27 | 3.709 |
| **13** | Dens_ComB | 0.259 | 3.867 | 0.259 | 3.858 | 0.262 | 3.824 |
| **14** | Dist_Res_CC | 0.231 | 4.33 | 0.259 | 3.858 | 0.281 | 3.56 |
| **15** | SLPrice | 0.227 | 4.405 | 0.228 | 4.39 | 0.229 | 4.363 |
| **16** | Dens_MetroExit | 0.17 | 5.88 | 0.173 | 5.764 | 0.181 | 5.523 |
| **17** | Dist_ ParkingLot | 0.163 | 6.153 | 0.163 | 6.142 | 0.165 | 6.076 |
| **18** | Dist_ Road_Sec | 0.161 | 6.205 | 0.166 | 6.017 | 0.186 | 5.381 |
| **19** | Dist_ Res_Qua | 0.16 | 6.264 | 0.165 | 6.045 | 0.165 | 6.05 |
| **20** | Dens_ParkingLot | 0.159 | 6.308 | 0.159 | 6.307 | 0.16 | 6.267 |
| **21** | Dens_Road | 0.148 | 6.743 | 0.148 | 6.735 | 0.152 | 6.573 |
| **22** | Dist_ Res_Dor | 0.145 | 6.887 | 0.146 | 6.855 | 0.142 | 7.035 |
| **23** | Dens_BusStop | 0.141 | 7.102 | 0.142 | 7.062 | 0.144 | 6.965 |
| **24** | Dens_ResB | 0.127 | 7.874 | 0.127 | 7.872 | 0.128 | 7.799 |
| **25** | Dist_Com_OffB | 0.103 | 9.718 | 0.121 | 8.286 | 0.157 | 6.383 |
| **26** | Dist_ MetroExit | 0.067 | 14.854 | 0.097 | 10.289 | Deleted | |
| **27** | Dist_ Com_ResB | 0.064 | 15.564 | Deleted | | | |

(c)

| | | **Central zone** | | **Middle zone** | | **Suburban zone** | |
|---|---|---|---|---|---|---|---|
| **No.** | **Variables** | **Tol** | **VIF** | **Tol** | **VIF** | **Tol** | **VIF** |
| **1** | POP | 0.8 | 1.3 | 0.5 | 2 | 0.3 | 3.4 |
| **2** | Dist_Waterarea | 0.7 | 1.4 | 0.7 | 1.3 | 0.7 | 1.4 |
| **3** | Dist_Road_Unc | 0.8 | 1.3 | 0.7 | 1.3 | 0.5 | 1.9 |
| **4** | Dist_Road_Spe | 0.6 | 1.6 | 0.4 | 2.3 | 0.4 | 2.7 |
| **5** | Slope | 0.6 | 1.7 | 0.6 | 1.7 | 0.5 | 2.2 |
| **6** | Dist_Road_Res | 0.6 | 1.6 | 0.5 | 1.8 | 0.4 | 2.3 |
| **7** | Dist_Road_Ter | 0.6 | 1.7 | 0.4 | 2.7 | 0.4 | 2.5 |
| **8** | Dist_Res_Vil | 0.6 | 1.7 | 0.5 | 2 | 0.4 | 2.5 |
| **9** | Dist_Road_Path | 0.5 | 1.9 | 0.4 | 2.5 | 0.3 | 3.2 |
| **10** | Dist_Res_Dor | 0.5 | 1.8 | 0.3 | 3.6 | 0.2 | 5 |
| **11** | Dist_BusStop | 0.4 | 2.2 | 0.3 | 3.5 | 0.4 | 2.4 |
| **12** | Dist_Res_CC | 0.5 | 2.2 | 0.3 | 3.2 | 0.4 | 2.8 |
| **13** | Dist_Road_Sec | 0.4 | 2.3 | 0.3 | 3.1 | 0.2 | 4.3 |
| **14** | DEM | 0.4 | 2.3 | 0.4 | 2.5 | 0.2 | 4.1 |
| **15** | Dist_Road_Pri | 0.4 | 2.5 | 0.7 | 1.4 | 0.4 | 2.3 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **16** | Dist_MetroExit | 0.4 | 2.5 | 0.4 | 2.8 | 0.1 | 6.9 |
| **17** | Dist_Com_OffB | 0.3 | 3.1 | 0.2 | 4.5 | 0.2 | 5.5 |
| **18** | Dist_Res_Qua | 0.3 | 3.3 | 0.2 | 5.9 | 0.1 | 8.3 |
| **19** | Dist_Com_ResB | 0.3 | 3.4 | 0.2 | 5.1 | 0.2 | 5.8 |
| **20** | SLPrice | 0.3 | 3.5 | 0.3 | 3.1 | 0.2 | 4.3 |
| **21** | Dist_ParkingLot | 0.3 | 4 | 0.2 | 6 | Deleted | |
| **22** | Dens_ComB | 0.2 | 5.9 | 0.3 | 2.9 | 0.1 | 7.8 |
| **23** | Dens_Road | 0.1 | 7.1 | 0.2 | 5.2 | 0.1 | 9.6 |
| **24** | Dens_BusStop | 0.2 | 4.8 | 0.2 | 6.4 | 0 | 0 |
| **25** | Dens_ResB | Deleted | | Deleted | | 0.1 | 7.9 |
| **26** | Dens_MetroExit | Deleted | | 0.3 | 3.3 | 0.3 | 3.6 |
| **27** | Dens_ParkingLot | Deleted | | 0.2 | 5.9 | 0.1 | 7.8 |

(d)

| | | **Central zone** | | **Middle zone** | | **Suburban zone** | |
|---|---|---|---|---|---|---|---|
| **No.** | Variables | Tolerance | VIF | Tolerance | VIF | Tolerance | VIF |
| **1** | POP | 0.6 | 1.6 | 0.5 | 2 | 0.3 | 3.7 |
| **2** | Dist_Waterarea | 0.7 | 1.4 | 0.7 | 1.4 | 0.5 | 1.9 |
| **3** | Dist_Road_Unc | 0.7 | 1.5 | 0.8 | 1.3 | 0.6 | 1.6 |
| **4** | Dist_Road_Spe | 0.7 | 1.4 | 0.6 | 1.8 | 0.4 | 2.8 |
| **5** | Slope | 0.7 | 1.4 | 0.7 | 1.4 | 0.5 | 1.8 |
| **6** | Dist_Road_Res | 0.5 | 1.9 | 0.5 | 2 | 0.3 | 3 |
| **7** | Dist_Road_Ter | 0.3 | 2.9 | 0.4 | 2.6 | 0.5 | 2 |
| **8** | Dist_Res_Vil | 0.6 | 1.6 | 0.5 | 2.2 | 0.4 | 2.5 |
| **9** | Dist_Road_Path | 0.6 | 1.7 | 0.4 | 2.4 | 0.3 | 3.2 |
| **10** | Dist_Res_Dor | 0.4 | 2.6 | 0.4 | 2.8 | 0.2 | 6.3 |
| **11** | Dist_BusStop | 0.3 | 3.9 | 0.3 | 3.3 | 0.5 | 2.1 |
| **12** | Dist_Res_CC | 0.4 | 2.5 | 0.3 | 3 | 0.3 | 3.5 |
| **13** | Dist_Road_Sec | 0.3 | 2.9 | 0.4 | 2.5 | 0.1 | 6.7 |
| **14** | DEM | 0.3 | 3.1 | 0.5 | 2.1 | 0.2 | 4.1 |
| **15** | Dist_Road_Pri | 0.5 | 2.1 | 0.7 | 1.4 | 0.3 | 3.6 |
| **16** | Dist_MetroExit | 0.3 | 2.9 | 0.4 | 2.5 | Deleted | |
| **17** | Dist_Com_OffB | 0.3 | 3.6 | 0.2 | 4.1 | 0.2 | 6.4 |
| **18** | Dist_Res_Qua | 0.3 | 3.9 | 0.2 | 5.8 | 0.2 | 5.7 |
| **19** | Dist_Com_ResB | 0.2 | 4.2 | 0.2 | 4.1 | Deleted | |
| **20** | SLPrice | 0.3 | 2.9 | 0.3 | 2.9 | 0.2 | 4.7 |
| **21** | Dist_ParkingLot | 0.2 | 5.2 | 0.2 | 4.7 | 0.2 | 5.8 |
| **22** | Dens_ComB | 0.2 | 4.5 | 0.3 | 3.1 | 0.2 | 6.2 |
| **23** | Dens_Road | 0.1 | 7.5 | 0.2 | 4.9 | 0.2 | 5.2 |
| **24** | Dens_BusStop | 0.2 | 5 | 0.2 | 5.3 | 0.1 | 12 |
| **25** | Dens_ResB | 0.1 | 7.8 | Deleted | | 0.1 | 9.5 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **26** | Dens_MetroExit | Deleted | 0.3 | 3.2 | 0.3 | 3.3 |
| **27** | Dens_ParkingLot | Deleted | 0.2 | 4.7 | 0.2 | 5.3 |

Table 3-6. Data of the outliers: (a) PPL and (b) PPS.

(a)

| **Case No.** | **Selected Status**[a] | **Observed** Y | **Predicted** | **Predicted Group** | **Temporary Variable** Resid | ZResid | SResid |
|---|---|---|---|---|---|---|---|
| **39** | S | 1** | 0.000 | 0 | 1.000 | 400.918 | 4.896 |
| **103** | S | 1** | 0.000 | 0 | 1.000 | 46.049 | 3.915 |
| **467** | S | 1** | 0.000 | 0 | 1.000 | 144.723 | 4.461 |
| **469** | S | 1** | 0.001 | 0 | 0.999 | 27.684 | 3.647 |
| **1144** | S | 0** | 0.991 | 1 | -0.991 | -10.574 | -3.084 |

(b)

| **Case No.** | **Selected Status**[a] | **Observed** Y | **Predicted** | **Predicted Group** | **Temporary Variable** Resid | ZResid | SResid |
|---|---|---|---|---|---|---|---|
| **15** | S | 1** | 0.000 | 0 | 1.000 | 1078.767 | 5.285 |
| **90** | S | 1** | 0.010 | 0 | 0.990 | 9.917 | 3.037 |
| **121** | S | 1** | 0.000 | 0 | 1.000 | 134.625 | 4.428 |
| **144** | S | 1** | 0.000 | 0 | 1.000 | 160.740 | 4.508 |
| **145** | S | 1** | 0.000 | 0 | 1.000 | 90.526 | 4.245 |
| **146** | S | 1** | 0.001 | 0 | 0.999 | 26.069 | 3.613 |
| **149** | S | 1** | 0.000 | 0 | 1.000 | 182.312 | 4.563 |
| **153** | S | 1** | 0.000 | 0 | 1.000 | 108.816 | 4.331 |
| **167** | S | 1** | 0.000 | 0 | 1.000 | 48.273 | 3.939 |
| **168** | S | 1** | 0.001 | 0 | 0.999 | 43.528 | 3.886 |
| **228** | S | 1** | 0.003 | 0 | 0.997 | 17.770 | 3.396 |
| **233** | S | 1** | 0.007 | 0 | 0.993 | 12.138 | 3.168 |
| **3395** | S | 0** | 0.991 | 1 | -0.991 | -10.219 | -3.055 |
| **3397** | S | 0** | 0.998 | 1 | -0.998 | -24.050 | -3.569 |
| **3866** | S | 0** | 0.996 | 1 | -0.996 | -15.592 | -3.317 |

### 3.1.4 Determination of the best explanatory variable combination

There is a large number of candidate explanatory variables in the model. It is important to detect the best variable combination for model fitting. A good model should adequately fit the data, and the predictor variables should not be too complicated. It is challenging to select the smallest number of candidate variables that can sufficiently predict the dependent variable while considering sample size constraints (Hosmer and Lemeshow, 1989). In previous studies using the LR model, forward and backward stepwise methods are often used (Zellner *et al*., 2004).

The forward stepwise selection method (FSSM) selects several significant predictor variables for the final model. Model optimization was performed using the least-squares criteria. It started with a blank model with no predictor variables. Variables were sequentially added one at a time to an empty model to predict the best output variable. Subsequently, a search for a second variable that can most improve the model fitting was conducted. The process was continued until a stopping rule was satisfied. In FSSM, variables added early in the process could be removed at a later stage because they became unimportant when other variables were added to the model. FSSM uses a systematic method for adding variables based on their statistical significance in a regression. The process starts with no explanatory variables in the model and then compares the incremental explanatory power of larger models (Soroush *et al*., 2012). Using the FSSM technique, a ranking list of the variable importance can be obtained according to the priority of the added variables.

Unlike FSSM, the backward stepwise elimination method (BSEM) starts with all predictors of the least-squares model and then eliminates the least effective predictor one at a time. This method continued until a stopping rule was satisfied. In the literature, the recommended stopping rule was a p-value of ~0.15 (Flack and Chang 1987; Lee and Koval 1997). In the SPSS software, the default values for FSSM and BSEM are 0.05 and 0.1, respectively.

### 3.1.5 Evaluation of the model performance

The performance comparison of the LR models is usually evaluated based on their discrimination and calibration. Discrimination refers to the ability of the model to correctly distinguish the two suitability classes based on prediction value. The discrimination capacity of the LR models is often measured by cross-classifying observations and predictions in the classification table and calculating indices of classification performance (Pearce and Ferrier, 2000). There are many indication metrics that can be used to measure the performance of a classifier or predictor. For example, sensitivity and specificity are often used in medicine, whereas precision and recall are preferred in computer science. However, precision and recall sometimes contradict each other. Therefore, it is necessary to find a comprehensive index to measure. The most common method is F-Measure (also known as F-Score). Here, precision, recall, accuracy and F-Measure would be the index of the discrimination.

$$\text{Precision} = \text{TP}/\ (\text{TP}+\text{FP}) \tag{3}$$

$$\text{Recall} = \text{TP}/\ (\text{TP}+\text{FN}) \tag{4}$$

$$\text{Accuracy} = (\text{TP}+\text{TN})/\ (\text{TP}+\text{FP}+\text{FN}+\text{TN}) \tag{5}$$

$$\text{F-Measure} = 2\times\text{Precision}\times\text{Recall}/\ (\text{Precision}+\text{Recall}) \tag{6}$$

TP, FP, TN, and FN denote true positive, false positive, true negative, and false negative, respectively.

Calibration describes how close the predicted value is to the actual value. The discrimination only compared the predicted probability value with a certain threshold; if it was below the threshold, the prediction result was “N” (0), otherwise, it was “Y” (1). The most common default threshold in binary LR is 0.5. Therefore, it ignores how far the predicted value is from the true value. Calibration resolves this shortcoming. The Brier score is an important calibration index that measures the accuracy of probabilistic predictions. It is applicable to tasks in which predictions assign probabilities to a set of mutually exclusive discrete outcomes. The set of possible outcomes can be either binary or categorical in nature, and the probabilities assigned to this set of outcomes must sum to 1, where each individual probability ranges from 0 to 1 (Brier, 1950). The lower the Brier score for a set of predictions, the better the predictions are calibrated.

$$B = \frac{\sum_{i=1}^{n}(x_i - q_i)^2}{n} \tag{7}$$

Here, x is the real dependent variable and q is the predicted probability.

In this study, $R^2$ was the model fitting degree, and the reduction ratio of the variables involved in modeling (model optimization rate) were added to evaluate the model performance.

### 3.2 Results and discussions

3.2.1 Optimum explanatory variable combination in standard LR model

Table 3-7 shows the model performance of FSSM and BSEM in the standard LR model. In the PPL, the accuracy and F-Measure of BSEM are higher than those of FSSM by 0.6% and 0.2%, respectively. The calibration and the model fitting of BSEM are also slightly better than FSSM. However, the optimization rate of FSSM is higher by 20%. In the PPS, the accuracy and F-Measure of BSEM are the same as those of FSSM, and the variable reduction ratio is < 8%. This indicates that the two methods of selecting the optimal variable combination in the standard LR model almost have the same result on the model fitting, accuracy, and bias. However, in the index of the model optimization, FSSM performed better than BSEM.

Table 3-8 shows the coefficient of the best explanatory variable combination in the standard LR model. The Wald value indicates the significance of variables. The coefficients in red represent the negative correlations. In the PPL data, eight significant variables without multicollinearity were selected from the 25 variables. One variable was selected from the seven types of road distance variables, three from the six types of building distance variables, one from the four density variables, one from the four transport variables, and two from other variables. According to the Wald value, the most crucial factor was Dist_Res_Qua with a value of 45.5, followed by SLPrice (29), Dist_BusStop (28.4), and Dens_ComBs (20.7). The highest Wald value was almost twice the second value, whereas the second and third values were almost the same. Based on the sign of the coefficient, the variables of Dist_Res_Quar, Dist_BusStop, Dist_Com_OffB, Dist_Road_Sec, Dist_Res_Vil, and SLPrice were negative correlated with the possibility of PPL in the mesh unit. DEM and Dens_ComB were positively correlated. Therefore, the PPL site may be situated close to residential quarters, commercial offices, or residential villas, with access to bus stops or secondary roads, and in high-density commercial buildings with relatively low land prices.

In the PPS data, 15 critical variables were selected from the 25 variables. Three variables were selected from the seven types of road distance variables: Dist_Road_Res, Dist_Road_Spe, and Dist_Road_Path. These three types of roads were commonly walked daily by residents. This indicated the residents collected parcels primarily by foot. Areas near residential roads, paths, and special types of roads need to be prioritized when selecting a site for PPS. Almost all variables without collinearity were selected from the five types of building distance variables except Dist_Res_CC. Thus, PPPs prefer areas close to different types of commercial or residential buildings that have a strong relationship with the daytime and nighttime populations. According to the Wald value, the most crucial factor was Dist_Res_Qua (210), followed by Dist_BusStop (114), and Dens_ComB (93). This implies that the PPS site can be located close to residential quarters, commercial offices, or residential dormitories, with good access to bus stops, residential road types, paths, or special road types, and far from parking lots and water. High-density commercial buildings, low-density metro exits, and bus stops with high population and relatively low land prices were the preferred locations.

The best variable combination of PPL and PPS data has several similarities as well as differences. The two PPP types have negative correlations with three factors: Dist_Res_Qua, Dist_Com_OffB, and Dist_BusStop, and a positive correlation with Dens_ComB. Residential buildings were more important than commercial buildings for PPSs, whereas the opposite was true for PPLs. For PPL, the proximity of the commercial buildings was more important than that of residential buildings. In addition, the types of dormitory buildings only have a significant impact on the location of PPS. Dormitories are typically present in schools or factory units, with a large floating population. The type of PPS was more suitable for the delivery needs of dormitories. There were more factors related to the transportation and road type variables for PPS. The PPS location was more accessible than the PPL location. The population variable only affected PPS. Here, the population corresponded to the nighttime population. The daytime population was more significant for PPL. The most significant difference between the two types of PPP was that commercial building variables have a greater impact on PPL. In comparison, residential buildings and dormitories have a greater impact on PPS. In addition, nighttime population and daily road factors of residents have a significant impact on PPS. This shows that PPS is suitable for site selection in densely populated areas with a convenient walking environment.

Table 3-7. Model performance of FSSM and BSEM in the standard LR model: (a) PPL and (b) PPS.

(a)

| | Discrimination | | | | Calibration | Optimization | Model fitting |
|---|---|---|---|---|---|---|---|
| **Method** | **Precision** | **Recall** | **Accuracy** | **F-Measure** | **Brier score** | **Reduction ratio** | $R^2$ |
| **FSSM** | 90.6% | 86.5% | 88.2% | 88.5% | 0.088 | 68.0% | 0.743 |
| **BSEM** | 91.2% | 86.4% | 88.4% | 88.7% | 0.085 | 48.0% | 0.755 |

(b)

| | Discrimination | | | | Calibration | Optimization | Model fitting |
|---|---|---|---|---|---|---|---|
| **Method** | **Precision** | **Recall** | **Accuracy** | **F-Measure** | **Brier score** | **Reduction ratio** | $R^2$ |
| **FSSM** | 90.8% | 83.7% | 85.4% | 87.1% | 0.01 | 40.0% | 0.677 |
| **BSEM** | 90.6% | 83.9% | 85.4% | 87.1% | 0.01 | 32.0% | 0.679 |

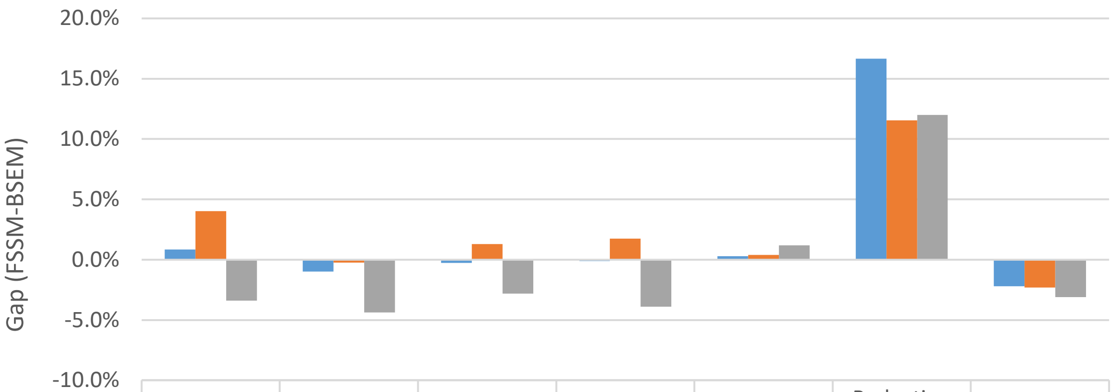
Table 3-8. Coefficient of the best explanatory variable combination in the standard LR model.

| Variable Type | Variable code | PPL | | PPS | |
|---|---|---|---|---|---|
| | | Coefficient | Wald | Coefficient | Wald |
| Others | DEM | 0.019 | 10.5 | | |
| | Slope | | | | |
| | POP | | | 0.0015 | 7.8 |
| | SLPrice | -0.0001 | 29 | -0.0001 | 36.8 |
| Building | Dist_Res_Qua | -0.0032 | 45.5 | -0.0028 | 210.6 |
| | Dist_Res_CC | | | | |
| | Dist_Res_Vil | -0.0002 | 6.7 | 0.0001 | 6.8 |
| | Dist_Res_Dor | | | -0.0004 | 28.1 |
| | Dist_Com_ResB | | | Collinearity | |
| | Dist_Com_OffB | -0.0013 | 18.6 | -0.0002 | 5.6 |
| Road | Dist_Road_Pri | | | | |
| | Dist_Road_Sec | -0.0006 | 9.3 | | |
| | Dist_Road_Ter | | | | |
| | Dist_Road_Unc | | | | |
| | Dist_Road_Res | | | -0.0004 | 10.1 |
| | Dist_Road_Spe | | | -0.0002 | 3.1 |
| | Dist_Road_Path | | | -0.0001 | 4.3 |
| Transport | Dist_MetroExit | | | Collinearity | |
| | Dist_BusStop | -0.0039 | 28.4 | -0.0029 | 114.0 |
| | Dist_ParkingLot | | | 0.0003 | 9.9 |
| | Dist_WaterArea | | | 0.0005 | 27.4 |
| Density | Dens_ParkingLot | | | | |
| | Dens_MetroExit | Collinearity | | -1.0479 | 31.5 |
| | Dens_BusStop | | | -0.2275 | 10.9 |
| | Dens_ComB | 0.090 | 20.7 | 0.1581 | 92.7 |
| | Dens_ResB | Collinearity | | | |
| | Dens_Road | | | | |

### 3.2.2 Best explanatory variable combination multi-zone LR model

#### *3.2.2.1 Performance of the variable selection methods of FSSM and BSEM*

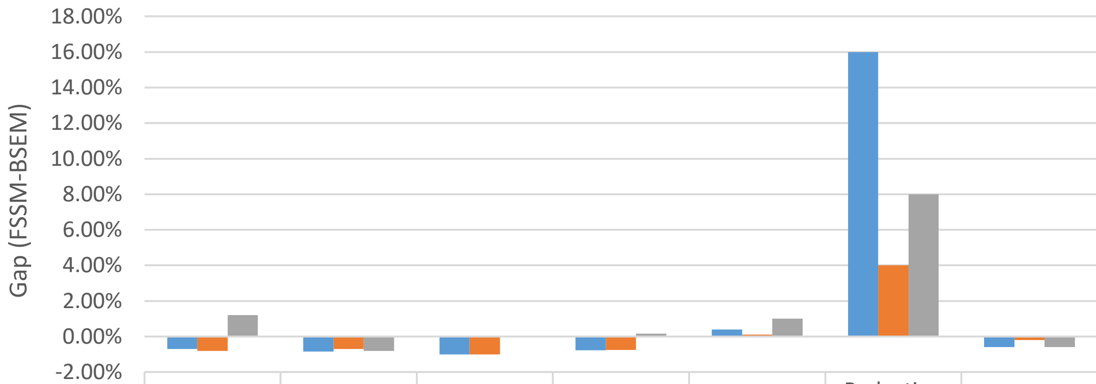
The performance indices between the two methods in the multi-zone model were compared. Figure 3-2 shows the gap index between FSSM and BSEM in the two types of PPPs. The blue, orange, and grey colors represent the central, middle, and suburban zones, respectively. For the PPL, the precision rate, accurate rate, F-Measure, and optimization rate of FSSM in the middle zone were larger than those of BSEM by 4%, 1.3%, 1.8%, and 11.5%, respectively. In the suburban zone, almost all the indices of FSSM were worse than BSEM. In the central zone, the optimization rate of FSSM was more significant than that of BSEM by 16%. The other index values in the two methods were quite similar. FSSM was better suited to the middle and central zones, and BSEM is the optimal method in the suburban zone. For the PPS, all the gap indices between the two methods were < 1% in the three zones except the model optimization. In model optimization, the reduction rate of FSSM was larger than that of BSEM, especially in the central zone. Hence, the classification accuracy of the two methods and bias under the multi-zone were not considerable. Therefore, FSSM yielded superior results to BSEM in the multi-zone model of PPS. Overall, FSSM had an absolute advantage in model optimization in that it minimized the number of variables in PPL and PPS. As BSEM has more variables involved in modeling than FSSM, it has advantages in model fitting and calibration. However, the gap was not highly significant.




| | Precision | Recall | Accuracy | F-Measure | Brier score | Reduction ratio | $R^2$ |
|---|---|---|---|---|---|---|---|
| Central zone | 0.9% | -1.0% | -0.3% | -0.1% | 0.003 | 16.7% | -0.022 |
| Middle zone | 4.0% | -0.2% | 1.3% | 1.8% | 0.004 | 11.5% | -0.023 |
| Suburban zone | -3.4% | -4.4% | -2.8% | -3.9% | 0.012 | 12.0% | -0.031 |

Central zone Middle zone Suburban zone

(a)




| | Precision | Recall | Accuracy | F-Measure | Brier score | Reduction ratio | $R^2$ |
|---|---|---|---|---|---|---|---|
| Central zone | -0.70% | -0.84% | -1.00% | -0.78% | 0.004 | 16.00% | -0.006 |
| Middle zone | -0.80% | -0.71% | -1.00% | -0.75% | 0.001 | 4.00% | -0.002 |
| Suburban zone | 1.20% | -0.80% | 0.00% | 0.16% | 0.01 | 8.00% | -0.006 |

**Central zone** **Middle zone** **Suburban zone**

(b)

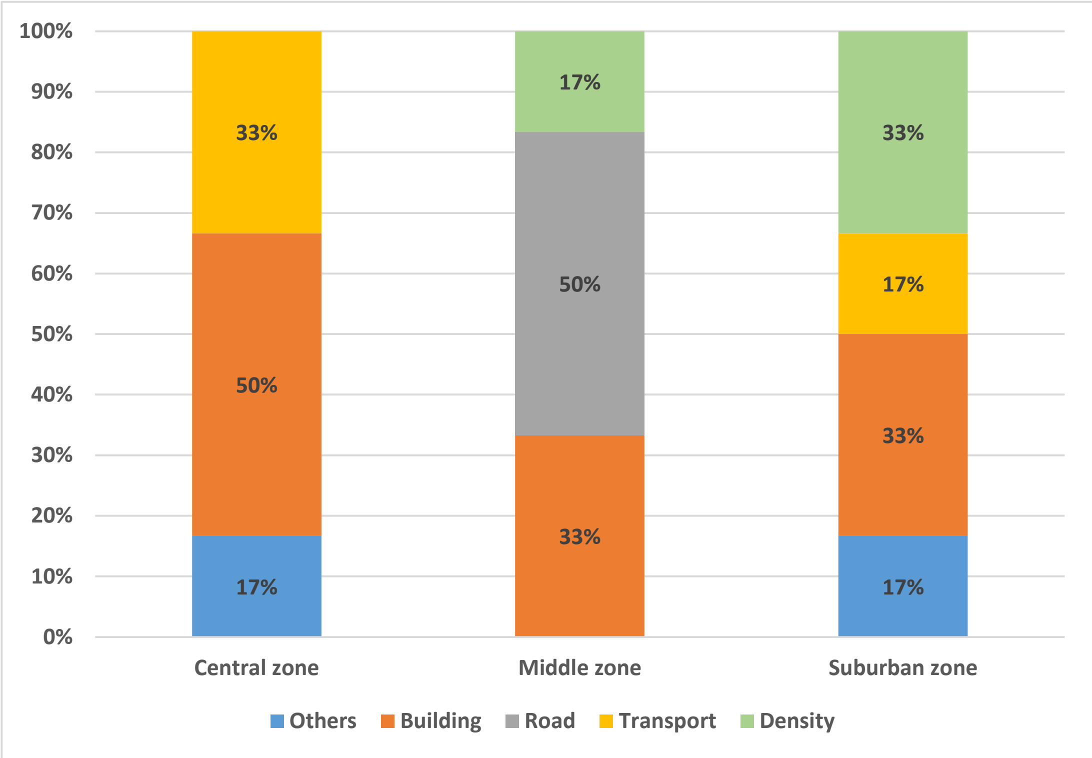
Figure 3-2. Performance gap between FSSM and BSEM in the multi-zone model (FSSM–BSEM): (a) PPL, and (b) PPS.

*3.2.2.2 Composition of the best variable combinations*

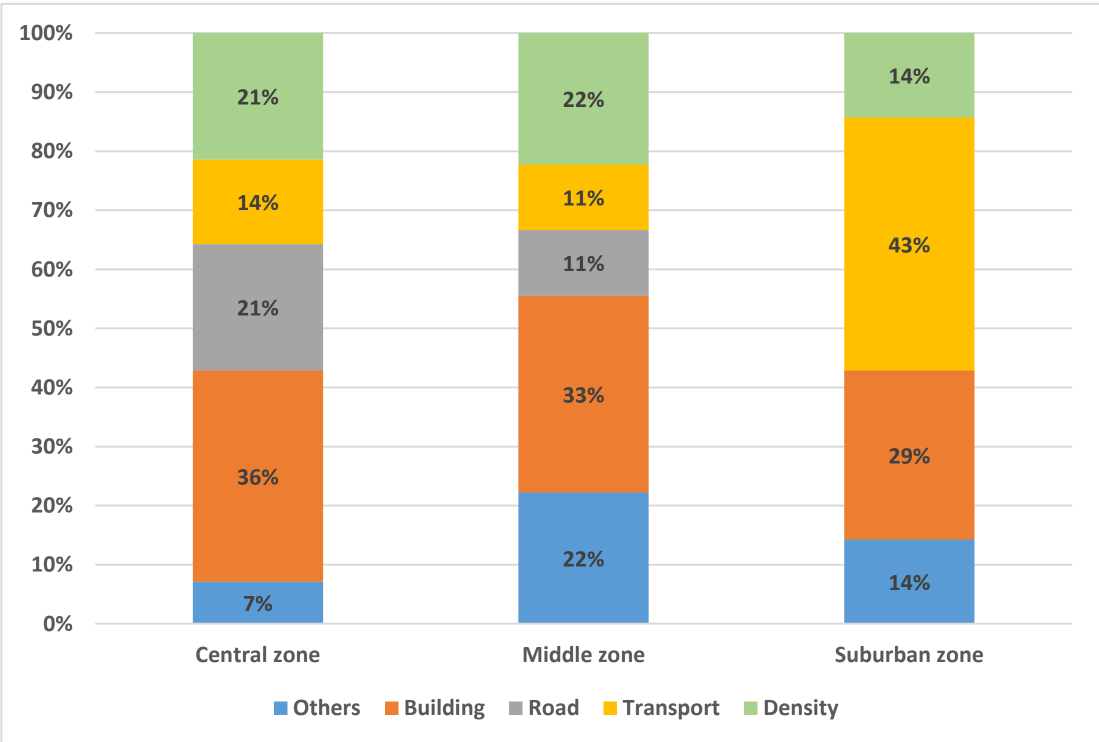
Figure 3-3 depicts the composition of the best variable combinations. In the PPL data, there were six optimal variables for the LR model in the three zones. Although the number of variables selected in these three zones were the same, the variables selected were entirely different. In the central zone, the variable of the distance from different types of buildings accounted for 50%; in the middle zone, the variable of the distance to various types of road accounted for 50%; in the suburban zone, various density variables and building distance variables both accounted for 33%. Various commercial areas and residential buildings were highly concentrated in the central zone. More variables of this type are expected. The metropolitan middle zone is the expansion of the central zone, and the traffic between them is significant. As the suburban zone is sparsely populated, more variables related to density and building distance were selected in this structure. For the PPS data, there were 14, 19, and 7 optimum variables for the LR models in the three metropolitan zones. The variable group of the distance to the type of building occupied the largest region in the central and middle zones, accounting for 36% and 33%. In the suburban zone, the largest part component was the transport group variables, which accounted for 33%.

The selected PPS and PPL variables in the multi-zone model revealed the same largest component in the central zone, but different components in the other two zones. In the middle zone, the largest components for the PPL and PPS were the distances to road types and building types, respectively. In the suburban zone, the largest component for the PPL was the various density and distance to the building types, whereas that for PPS was the distance to transportation. This indicated that the largest variable composition categories of the distribution factors for PPL and PPS differed under the different metropolitan structure zones. Even in the same metropolitan structure zone, the largest variable composition categories for the PPL and PPS suitability models differed.

.




(a) PPL




(b) PPS

Figure 3-3. Composition of the best variable combinations in the three structures: (a) PPL, and (b) PPS.

*3.2.2.3 Coefficient of the best explanatory variable combination*

Table 3-9 shows the coefficient of the best explanatory variable combination in the multi-zone LR model for PPL. The Wald value indicated the significance of the variables. In the central zone, the most crucial factor was Dist_Res_Qua (Wald value: 17.4), followed by Dist_Com_ResB (Wald value: 11.75), and Dist_ParkingLot (Wald value: 7.72). In the middle zone, the most crucial factor was Dist_Road_Ter (Wald value: 9.48), followed by Dens_ComB (Wald value: 8.11). In the suburban zone, the most crucial variable was Dist_BusStop (Wald value: 6.8), followed by Dist_Res_Qua (Wald value: 6.12). PPLs in the central zone were mainly located in residential areas and near parking lots. In the middle zone, they were mainly located near the third road type. In the suburban zone, they were mainly located near the bus station. Further, in these three metropolitan structure zones, Dist_Res_Qua was the only factor with the same selected variable.

Table 3-10 shows the coefficient of the best explanatory variable combination in the multi-zone LR model for PPS. In the central zone, the most crucial factor was Dist_Res_Qua (Wald value: 97.45), followed by Dist_WaterArea and SLPrice. In the middle zone, the most crucial factor was Dens_ComB (Wald value: 40.49), followed by SLPrice and Dist_BusStop. In the suburban zone, the most crucial variable was Dist_BusStop (Wald value: 33.12), followed by Dist_Res_Qua and Dist_Res_Dor. Suitable PPS locations in the central zone include the surrounding areas of residential, dormitory, and commercial office buildings. A convenient walking environment was located close to the residential road types and bus stations, away from water areas, low road density, and high density of commercial buildings. In the middle zone, the suitable PPS location was in the surrounding areas of the residential quarters and dormitories, close to special road types and bus stops, with a high density of commercial buildings. In the suburban zone, the suitable area was close to residential quarters, dormitories, and bus stations, and had a high-density road network and population.

In the PPL and PPS multi-zone models, the most critical factors in the central and suburban zones were the same, namely Dist_Res_Qua and Dist_BusStop. The critical factors were different in the middle zone. The most crucial variable for the PPL was Dist_Road_Ter, and there were no transport variables selected in this zone. This indicates that the primary means of collecting parcels from PPLs in the middle zone may be by driving, which is closely related to this road type. For the PPS, Dist_Res_Qua, Dist_Res_Dor, and Dist_BusStop were the same variables selected in the three metropolitan structure zones. Dist_Com_OffB was only selected in the central zone, which indicates that PPSs mainly target residential or dormitory residents. The suitable areas for PPS were near residential buildings and bus stops. The consumers may collect parcels on their way home.

Table 3-9. Coefficient of the best explanatory variable combination in multi-zone LR model for PPL.

| Variable Type | Variable code | Central zone | | Middle zone | | Suburban zone | |
|---|---|---|---|---|---|---|---|
| | | Coefficient | Wald | Coefficient | Wald | Coefficient | Wald |
| **Others** | DEM | | | | | | |
| | Slope | 0.104 | 6.92 | | | | |
| | POP | | | | | 0.025 | 3.4 |
| | SLPrice | | | | | | |
| **Building** | Dist_Res_Qua | -0.0060 | 17.4 | -0.001 | 2.83 | -0.0020 | 6.12 |
| | Dist_Res_CC | | | | | | |
| | Dist_Res_Vil | -0.0002 | 4.78 | | | | |
| | Dist_Res_Dor | | | | | | |
| | Dist_Com_ResB | -0.0011 | 11.75 | | | | |
| | Dist_Com_OffB | | | -0.0010 | 4.92 | -0.0010 | 3.42 |
| **Road** | Dist_Road_Pri | | | | | | |
| | Dist_Road_Sec | | | -0.0010 | 5.4 | | |
| | Dist_Road_Ter | | | -0.0020 | 9.48 | | |
| | Dist_Road_Unc | | | | | | |
| | Dist_Road_Res | | | 0.0010 | 4.69 | | |
| | Dist_Road_Spe | | | | | | |
| | Dist_Road_Path | | | | | | |
| **Transport** | Dist_MetroExit | | | | | | |
| | Dist_BusStop | -0.0030 | 5.75 | | | -0.0040 | 6.5 |
| | Dist_ParkingLot | -0.0046 | 7.72 | | | Collinearity | |
| | Dist_WaterArea | | | | | | |
| **Density** | Dens_ParkingLo | Collinearity | | | | 0.198 | 2.5 |
| | Dens_MetroExit | Collinearity | | | | | |
| | Dens_BusStop | | | | | | |
| | Dens_ComB | | | 0.1470 | 8.11 | | |
| | Dens_ResB | Collinearity | | Collinearity | | -0.1950 | 2.78 |
| | Dens_Road | | | | | | |

Table 3-10. Coefficient of the best explanatory variable combination in multi-zone LR model for PPS.

| Variable Type | Variable code | Central zone Coefficient | Central zone Wald | Middle zone Coefficient | Middle zone Wald | Suburban zone Coefficient | Suburban zone Wald |
|---|---|---|---|---|---|---|---|
| Others | DEM | | | | | | |
| | Slope | | | | | | |
| | POP | | | 0.0050 | 4.15 | 0.0124 | 4.63 |
| | SLPrice | -0.0001 | 21.16 | -0.0001 | 24.11 | | |
| Building | Dist_Res_Qua | -0.0058 | 97.45 | -0.0007 | 15.33 | -0.0008 | 16.49 |
| | Dist_Res_CC | -0.0006 | 12.75 | | | | |
| | Dist_Res_Vil | 0.0002 | 8.79 | | | | |
| | Dist_Res_Dor | -0.0006 | 4.02 | -0.0003 | 9.95 | -0.0003 | 13.69 |
| | Dist_Com_ResB | | | 0.0001 | 6.07 | | |
| | Dist_Com_OffB | -0.0014 | 13.53 | | | | |
| Road | Dist_Road_Pri | 0.0004 | 5.96 | | | | |
| | Dist_Road_Sec | | | | | | |
| | Dist_Road_Ter | | | | | | |
| | Dist_Road_Unc | -0.0012 | 7.47 | | | | |
| | Dist_Road_Res | -0.0015 | 9.38 | | | | |
| | Dist_Road_Spe | | | -0.0007 | 14.36 | | |
| | Dist_Road_Path | | | | | | |
| Transport | Dist_MetroExit | | | | | | |
| | Dist_BusStop | -0.0027 | 16.05 | -0.0012 | 19.42 | -0.0034 | 33.12 |
| | Dist_ParkingLot | | | | | 0.0002 | 5.34 |
| | Dist_WaterArea | 0.0019 | 29.31 | | | 0.0003 | 3.87 |
| Density | Dens_ParkingLot | Collinearity | | | | | |
| | Dens_MetroExit | Collinearity | | | | | |
| | Dens_BusStop | -0.3341 | 8.47 | | | Collinearity | |
| | Dens_ComB | 0.0755 | 14.46 | 0.3326 | 40.49 | | |
| | Dens_ResB | | | Collinearity | | | |
| | Dens_Road | -0.1173 | 7.8 | 0.134 | 7.91 | 0.2311 | 10.55 |

### 3.2.3 Model performance of the standard and multi-zone LR models

The test dataset was used to conduct an unbiased evaluation of the final model fit on the training dataset. The final standard and multi-zone LR models with the best variable combination and their coefficients were applied to the PPL and PPS test dataset. The F-measure and Brier score were the indicators that evaluated the performance of the two models.

Figure 3-4 shows the predicted performance of the final multi-zone and standard LR models in the test dataset. The blue bar represents the performance in the different zones of the multi-zone model, and the orange line represents the performance of the standard model. The larger the F-measure index, the higher the discrimination accuracy of the model's classification. The F-Measure values of PPL in the multi-zone model were all greater than 90%, and that of PPS were greater than 79%. The middle zone had the highest F-Measure index. Only the central zone was slightly worse than the standard model. The other two zones and the average of all the zones were better than the standard model. The lower the Brier score, the smaller the deviation predicted, and the higher the calibration degree of the model. The Brier score of the two PPP types was quite low, and the value for PPL was smaller than that for PPS. The smallest values for the PPL and PPS were in the middle and suburban zones, respectively. The average Brier scores in all the zones in the two PPP types were lower than those in the standard model. Overall, the predicted performance of the multi-zone LR model in the test dataset was better than the standard LR model.

To compare the final bias of the multi-zone and standard LR models, the final two models were applied to classify the suitability throughout the entire study area. Figures 3-5 (a) and (b) show the suitable areas predicted for PPLs and PPSs in Guangzhou by the two models. The blue and orange bars represent the standard and multi-zone LR models, respectively. The predicted suitable areas for PPS were larger than the PPL. The suitable areas in Guangzhou for the PPL and PPS predicted by the multi-zone LR model were 743 sq. km and 1148 sq. km, respectively. The suitable size of PPS was ~1.5 times larger than that of PPL. Figure 3-5 (c) shows the gap between the two models of PPL and PPS. The gap of PPL was larger than that of PPS, particularly in the middle and suburban zones. The predicted area of the multi-zone model was only smaller than the standard model in the central zone. Other zones and the whole area of the multi-zone model were all larger than the standard model. This indicates that the overall size of the prediction area will be reduced using the standard LR model, which has a lower performance than the multi-zone model, except for the expected value in the central area.

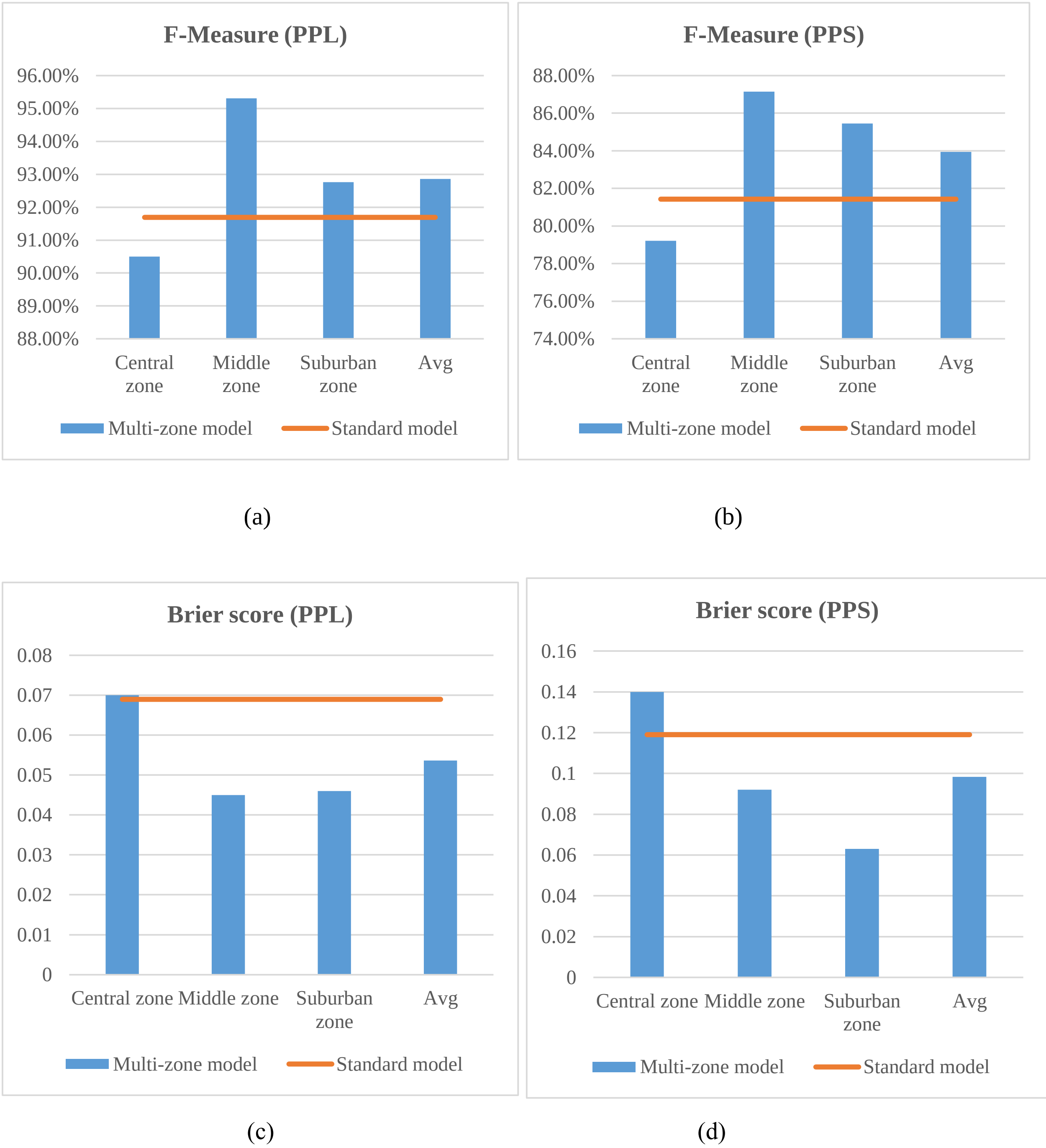
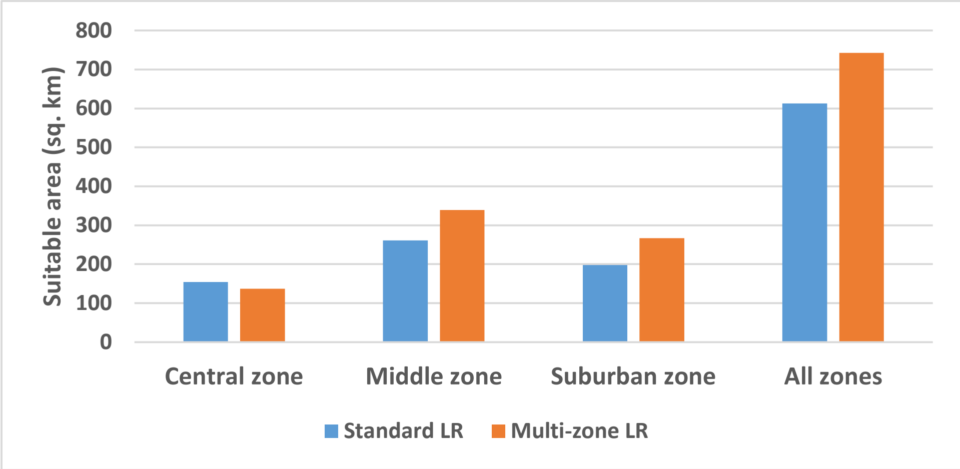
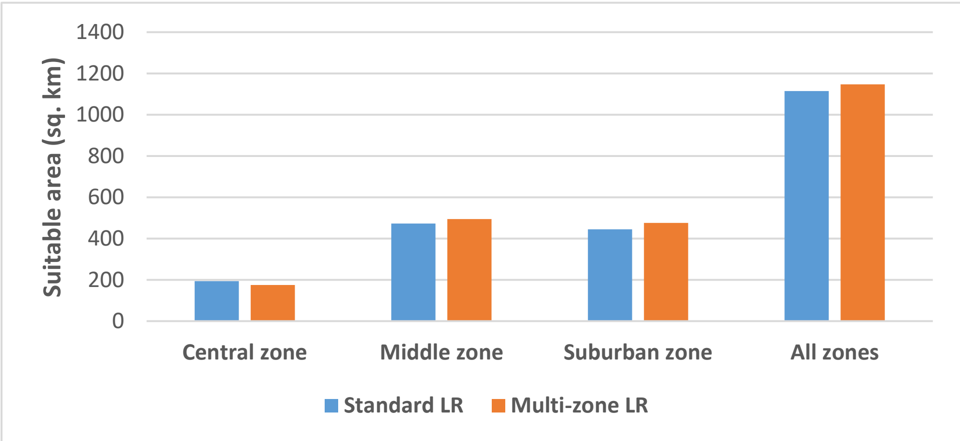


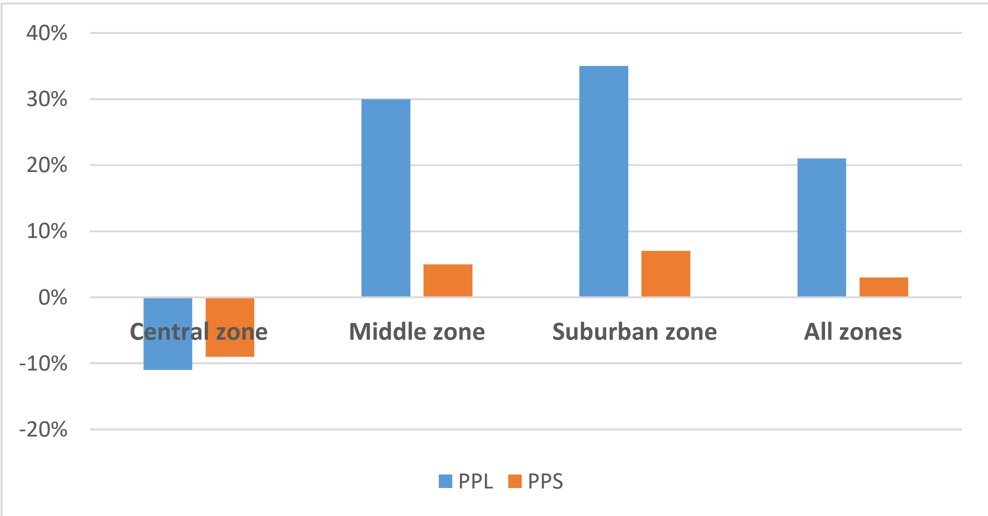

Figure 3-4. Predicted performance of the final multi-zone and standard LR models in the test dataset: (a) F-measure index of PPL, (b) F-measure index of PPS, (c) Brier score of PPL, and (d) Brier score of PPS.




(a) PPL




(b) PPS




(c) Gap%: (Multi-zone LR–Standard LR)/Standard LR

Figure 3-5. Suitable area predicted by the two models in Guangzhou: (a) PPL, (b) PPS. (c) gap between two models of PPL and PPS.

### 3.2.4 Suitable area results by multi-zone LR model

Figure 3-6 and Figure 3-7 demonstrate the results of suitable areas for PPL and PPS simulated by a multi-zone LR model. Overall, the size of the suitable area for PPS was more than that of PPL. In the central zone, the suitable area for PPS was mostly distributed over most of the area. PPL was concentrated in the middle, and there were almost no suitable areas in the northern, southwest, and southeast areas. In the middle zone, the most suitable areas for the two types of PPPs were close to the central zone, and there was almost no suitable area in the northeast region. In the suburban zone, suitable areas were dispersed in some small areas, and most of the suburban zone was not suitable for PPPs.

Figure 3-8 shows the summary of suitable areas for two types of PPPs in the three structure zones. The middle zone has the largest suitable area, followed by the suburban and central zones. The suitable areas in the middle zone and the suburban area were almost the same and were almost four times that of the central zone. According to the proportion of the three metropolitan structure zones, the largest proportion was in the central zone for PPS, accounting for 54.5%. The suburban zone was the smallest, accounting for only 5.4%. Thus, as the degree of urbanization gradually decreases from the metropolitan central zone to the suburban zone, the proportion of PPP suitable areas gradually becomes small. Overall, the PPS suitable areas account for 16.5% of Guangzhou's total area, while PPL accounts for 10.7 %. The selection range of PPPs can be focused on these suitable areas, significantly reducing the analysis difficulty and time cost.

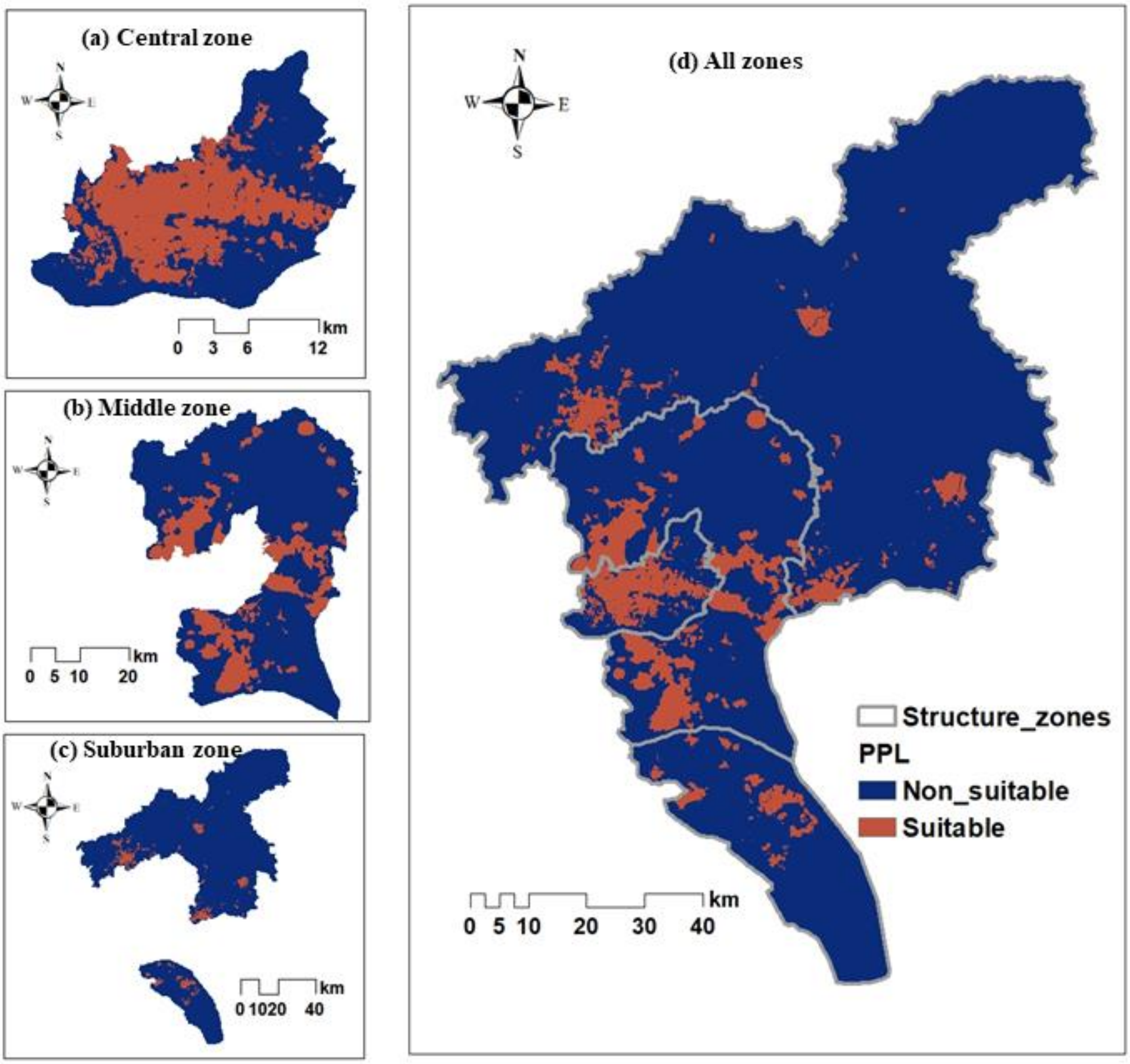



Figure 3-6. Suitability map for PPL using a multi-zone LR model: (a) central zone, (b) middle zone, (c) suburban zone, and (d) all zones.

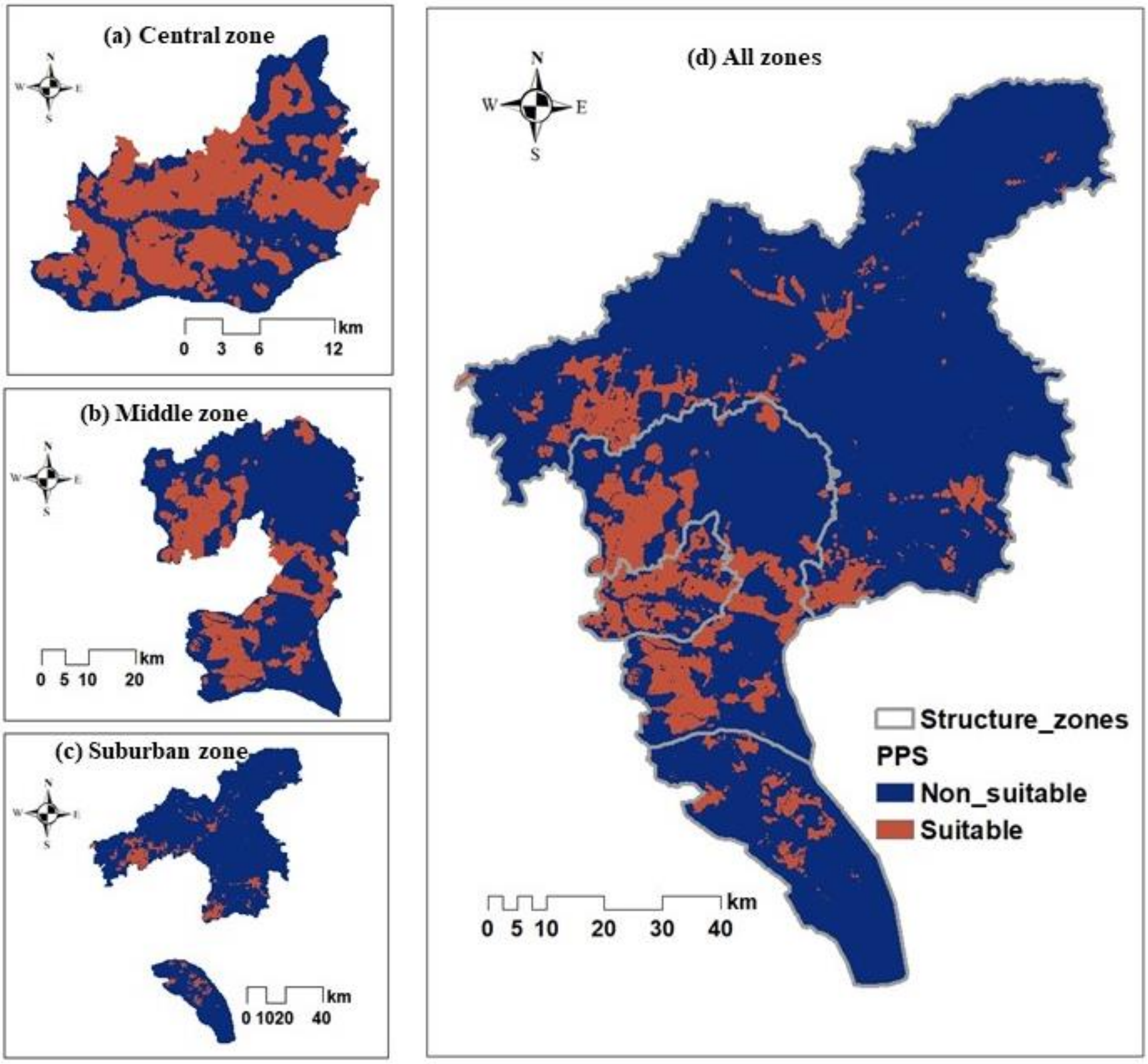



Figure 3-7. Suitability map for PPS using a multi-zone LR model: (a) central zone, (b) middle zone, (c) suburban zone, and (d) all zones.

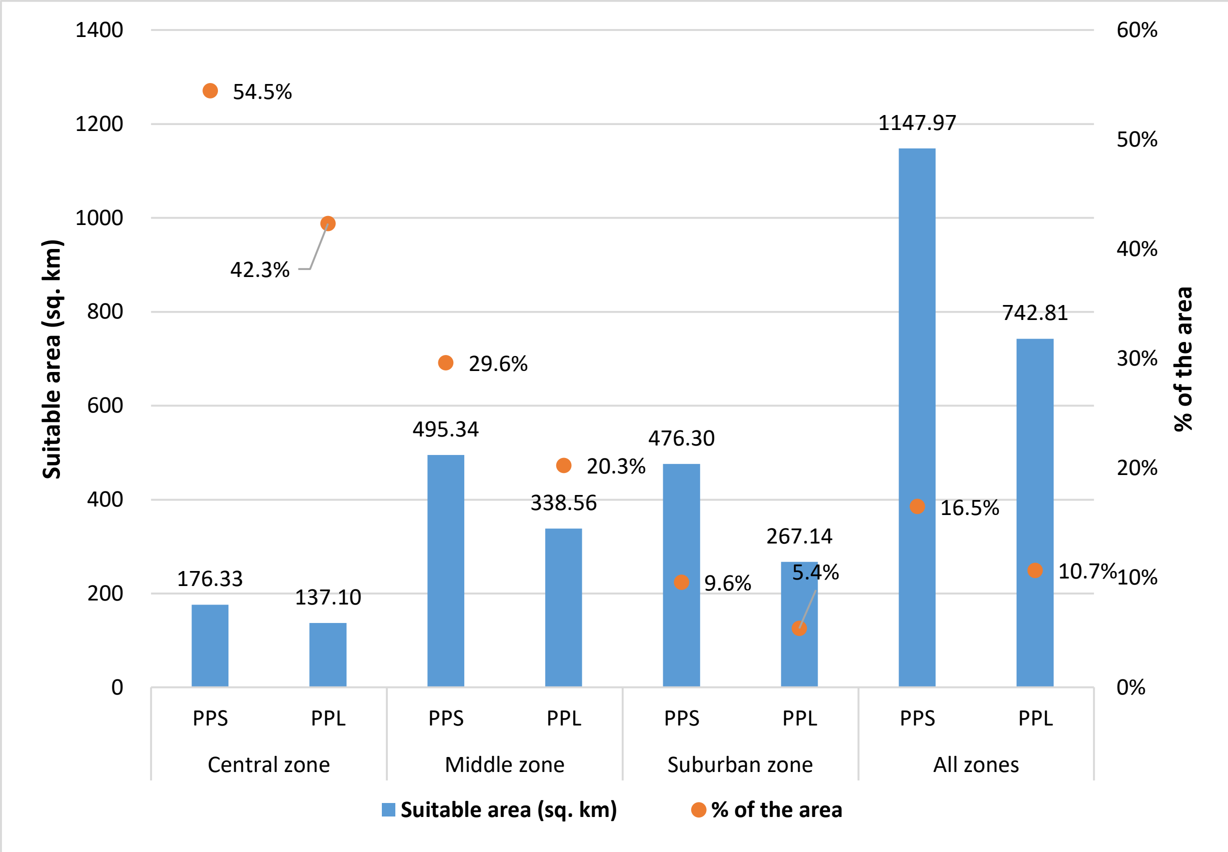



Figure 3-8. Summary of the suitable area for PPS and PPL using the multi-zone LR model.

### 3.3 Summary

This chapter aims to explore the relationship between the two types of PPP and surrounding geographic factors based on the simulation of the suitability using the LR model of ML.

In addition to defining the research unit as the smallest pixel, the variables that affect the distribution were also divided to enable microscopic analysis. The roads and building factors were refined into seven road types and six building categories to explore the specific types of roads or buildings with the most significant effect on their site selection and the consistency of the variables affecting the PPL and PPS. A total of 27 influencing variables were collected for analysis. This chapter compares the prediction results of the standard LR model with those of the multi-zone LR model to evaluate the model that performed better. In addition, owing to the many candidate variables, FSSM and BSEM methods were used to optimize the candidate variables to determine the best combination of model calculation variables, speeding up the computation and reducing the difficulty.

The results of the best combination of variables generated by FSSM and BSEM revealed that the results had little difference in terms of the model accuracy. However, FSSM was better than BSEM in optimizing the number of variables. The prediction results and observations of the multi-zone and standard LR models revealed that the accuracy and deviation rate of the multi-zone LR model were higher overall, but the accuracy of the central zone was slightly lower. Judging from the area of the suitability area in Guangzhou predicted by the two models, the region generated by the multi-zone LR model was generally larger; the area in the central zone was slightly smaller. The most important variable affecting the distribution of PPL and PPS in the standard LR model was Dist_Res_Qua. In the multi-zone LR model, the most important variables in the central and suburban zones were Dist_Res_Qua and Dist_BusStop, respectively. The most important variable of PPL and PPS in the middle zone was different.

In general, the results obtained using the standard LR model with one structure were sometimes difficult to interpret, and the symbols of the variables were not logical. However, results obtained using the multi-zone LR models were simpler and reflected the characteristics of different structures. The variable selection was more reasonable, and the degree of explanation was higher. Therefore, the performance of the multi-zone model was better than that using only one structure. The scope of the generated suitability area is more reliable. The suitable area for PPS was considerably larger than that for PPL, and the area gradually decreased from the central zone to the suburban zone.

# Chapter 4 Spatial relationship analysis of the suitable areas for PPPs

The previous chapter simulated the suitable areas for the location selection of PPPs in the whole Guangzhou metropolitan area. Using the results of Chapter 3, this chapter aims to investigate the relationship of the suitable areas of the two types of PPPs and their interaction. It explores the coexisting relationship of PPPs to avoid competition and can provide decision-makers with insight into how to position PPPs in suitable areas for PPP sustainability.

As the two types of PPPs, PPL and PPS share similar resources and may compete with each other to an extent. However, there are no reports on the competition between the two types of PPPs or the region for competition. The niche overlap theory can explain how competing species interact and coexist in environments with limited resources. This chapter aims to apply the niche overlap theory to geographical location to analyze the interaction between two types of PPPs. In ecology, each species occupies its niche in an ecosystem. The niche width of one species is based on the surrounding environment and resource. When more than one species exists, their niche relationship has three cases: complete separate, complete overlap, and partial overlap. The partial overlap is the most prominent. In the niche overlap area, the species share resources or compete with one another. Research on the ecological niche can help reduce competition from other species. This theory has been applied in economics, such as niche markets. This study aims to reduce the competition between PPS and PPL on the basis of this theory.

Figure 4-1 shows the different types of competition experienced by PPPs in the niche overlap theory. There are two types of competition in the overlapping area are: intertype and intratype. In the non-overlapping area, only intratype competition exists. As there is no competition from other types, this type of PPPs may survive. However, the same type of PPPs cannot reproduce indefinitely. To a certain extent, intratype competition will inevitably occur. Therefore, this study analyzes the appropriate spacing between individuals of the same type or different types in the suitable areas for PPPs to avoid intra- and inter-species competition and achieve coexistence or cooperation.

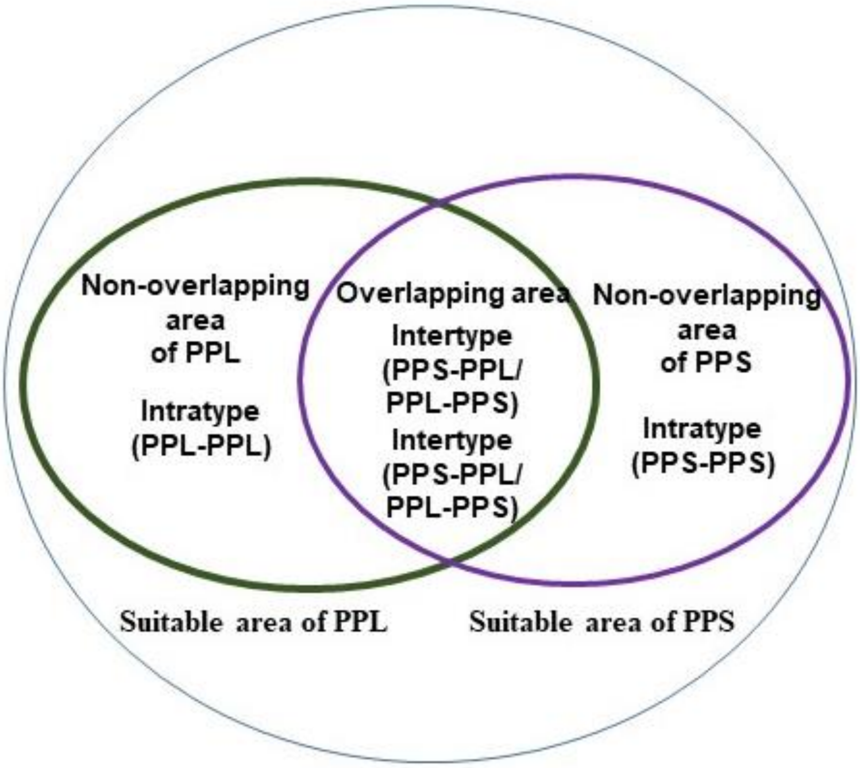



Figure 4-1. Different types of competitions based on the niche overlap theory.

## 4.1 Methodology

The methodology in Figure 4-2 is mainly divided into three steps:

(1) The relationship between the suitable areas for PPLs and PPSs was analyzed using the overlay method, and the degree of overlap on the different metropolitan structure zones was explored.

(2) In the overlapping area, both intertype and intratype competitions exist. As the two types of PPPs have been established in Guangzhou for several years, the unsuitable locations were closed. The existing locations were tested by the market and could coexist with each other. Hence, the coexisting relationship was analyzed based on the distribution of the existing PPLs and PPSs. The coexisting relationship was analyzed based on the following: the number of coexisting PPPs within a certain distance and the distance of the nearest coexisting PPP. High compatibility indicated more coexisting PPPs in the surrounding and a short distance to the nearest coexisting PPP. The appropriate spacing of the two PPP types in the overlap area can avoid their competition.

(3) In the non-overlapping area, intraspecific competition only existed theoretically. The same method in step 2 was used to explore the appropriate degree of reproduction in this area to avoid overexploitation.

Distance analysis of all the facilities was conducted by ArcMap software. The distance to the nearest facility was collected for each point to calculate the average distance and determine the intertype or intratype competition of PPPs. Survey results of resident preferences in Guangzhou (Zheng *et al*., 2020) revealed that most residents selected the following three acceptable pickup distances: within 100 m, 300 m, and 500 m. These three distance parameters were used to analyze the number of coexisting points around a facility. Figure 4-3 shows a sample calculation of the number of intertype and intratype coexisting points for one PPL at the three distances. The average number of intertype and intratype coexisting points of all individuals and the percentage of PPPs with coexisting points were calculated. These indicators were used to detect the compatibility of the two types of PPP.

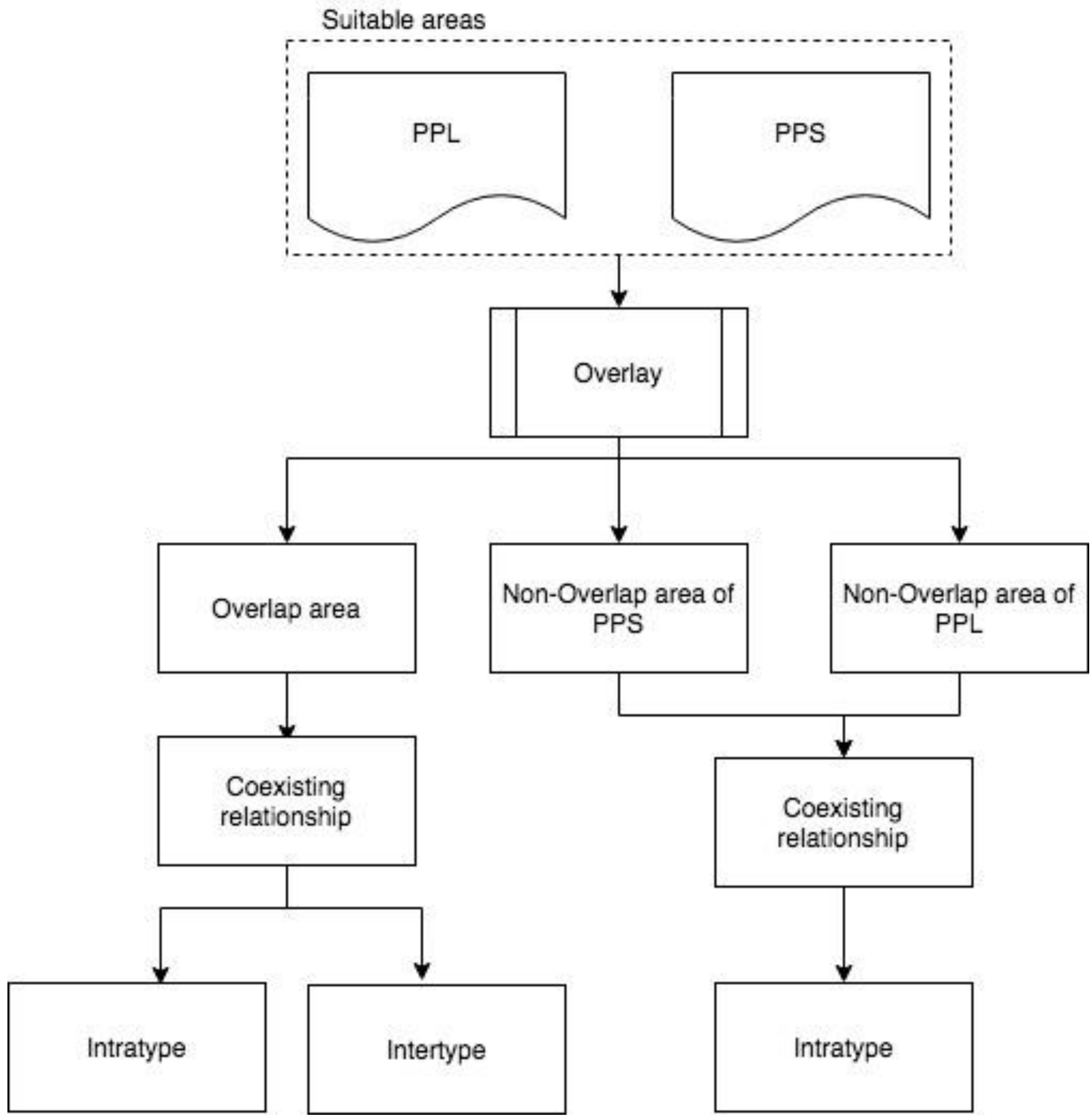
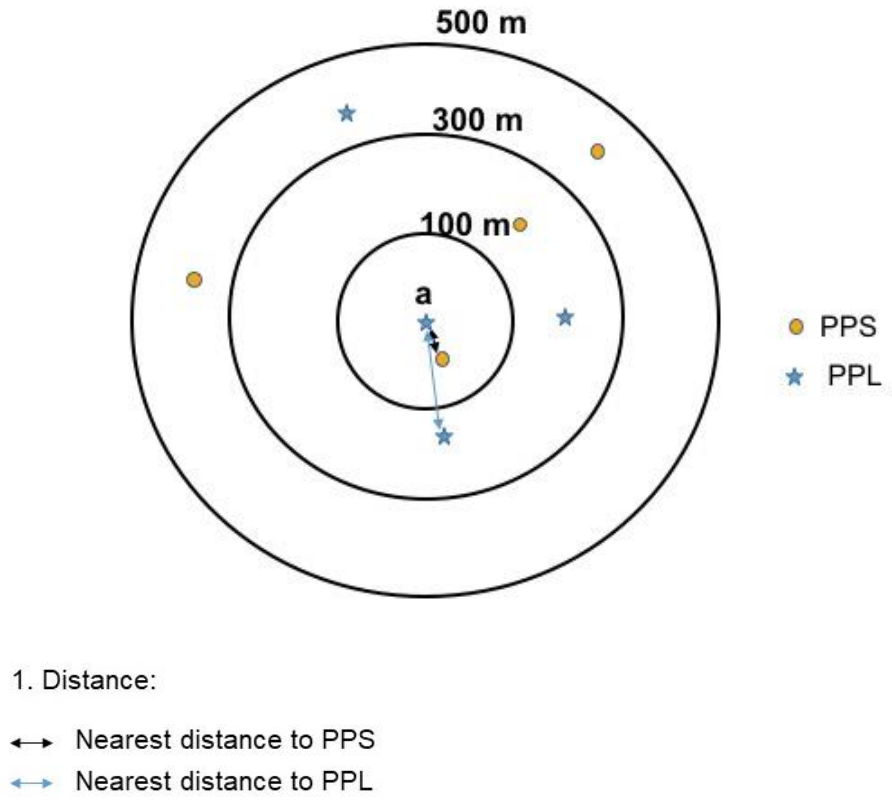



Figure 4-2. Methodology used in Chapter 4.




2. Number of coexisting points

| Number of coexisting points | Distance | | |
|---|---|---|---|
| | 100 m | 300 m | 500 m |
| **Intertype (PPL-PPS)** | 1 | 2 | 4 |
| **Intratype (PPL-PPL)** | 0 | 2 | 3 |

Figure 4-3. Sample calculation of the compatibility indicators for one PPL point.

### 4.2 Relationship between the suitable area for the PPP types

In this section, overlay analysis is performed in the suitable areas for PPLs and PPSs. Figure 4-4 (a), (b). and (c) shows the positional relationship between the suitable areas for PPL and PPS in the different metropolitan structure zones. There are three types of relationship labels. The green, red, and blue areas represent the overlap areas for PPL and PPS, the non-overlap area for PPL, and the non-overlap area for PPS. In the central zone, the non-overlap area for PPL is distributed around the overlap area and clustered near the border of the four administrative districts. The non-overlap area for PPS is an extension of the overlap area, distribution away from the overlap area. For example, several small areas in the east and north of the upper-left Tianhe district, south of the lower-left Haizhu district, and the west of the upper-left Liwan district are isolated. In the middle zone, the non-overlap area for PPL is scattered away from the overlap, and the non-overlap area for PPS is an extension of the overlap area. In the suburban zone, the non-overlap area for PPL is extremely small. The non-overlap area for PPS is mainly distributed in the Huadu, Qingyuan, and Conghua districts, with less distribution in the Nansha district. The proportions of the three labels in different structure zones are shown in Figure 4-4(d). Overall, the percentage of the overlapping area exceeded 50%, which is larger than that of the two non-overlap areas. The percentage of the non-overlap area is the least for PPL, only accounting for < 10%, whereas the percentage of the non-overlap area for PPS is relatively large (~30–50%). In the three structure zones, the largest percentage of the overlap area is in the central zone (~60%), followed by the middle and suburban zones. The largest and smallest percentages of the non-overlap area for PPL are in the central (11%) and suburban (4%) zones, respectively. These two relationship labels exhibit a trend of linear decrease from the central to the suburban zones. The largest proportion of non-overlap area for PPS is in the suburban zone (46%), whereas the smallest is in the central zone. The trend observed was opposite to that of the two previous relationship labels.

The relationship between the suitable areas for PPLs and PPSs is generally a partial overlap; the size of the overlapping area is rather massive. The suitable areas for PPL, particularly in the suburban zone, are almost included in the suitable areas for PPS. The proportion of the overlap and non-overlap areas for PPL decreases from the central zone to the suburban zone. The proportion of the non-overlap area for PPS exhibits an opposite trend—increasing from the central zone to the suburban zone. Compared to the suburban zone, the central and middle zones are more urbanized, with a developed economy, convenient transportation, and dense population. The non-overlap area for PPS is larger than that for PPL. The non-overlap area is only occupied by one type, and there was no competition between the different types of PPPs. Therefore, the competitive superiority of PPS is stronger than that of PPL, particularly in the suburban zone. The competition between the two PPP types gradually decreases from the central zone to the suburban zone.

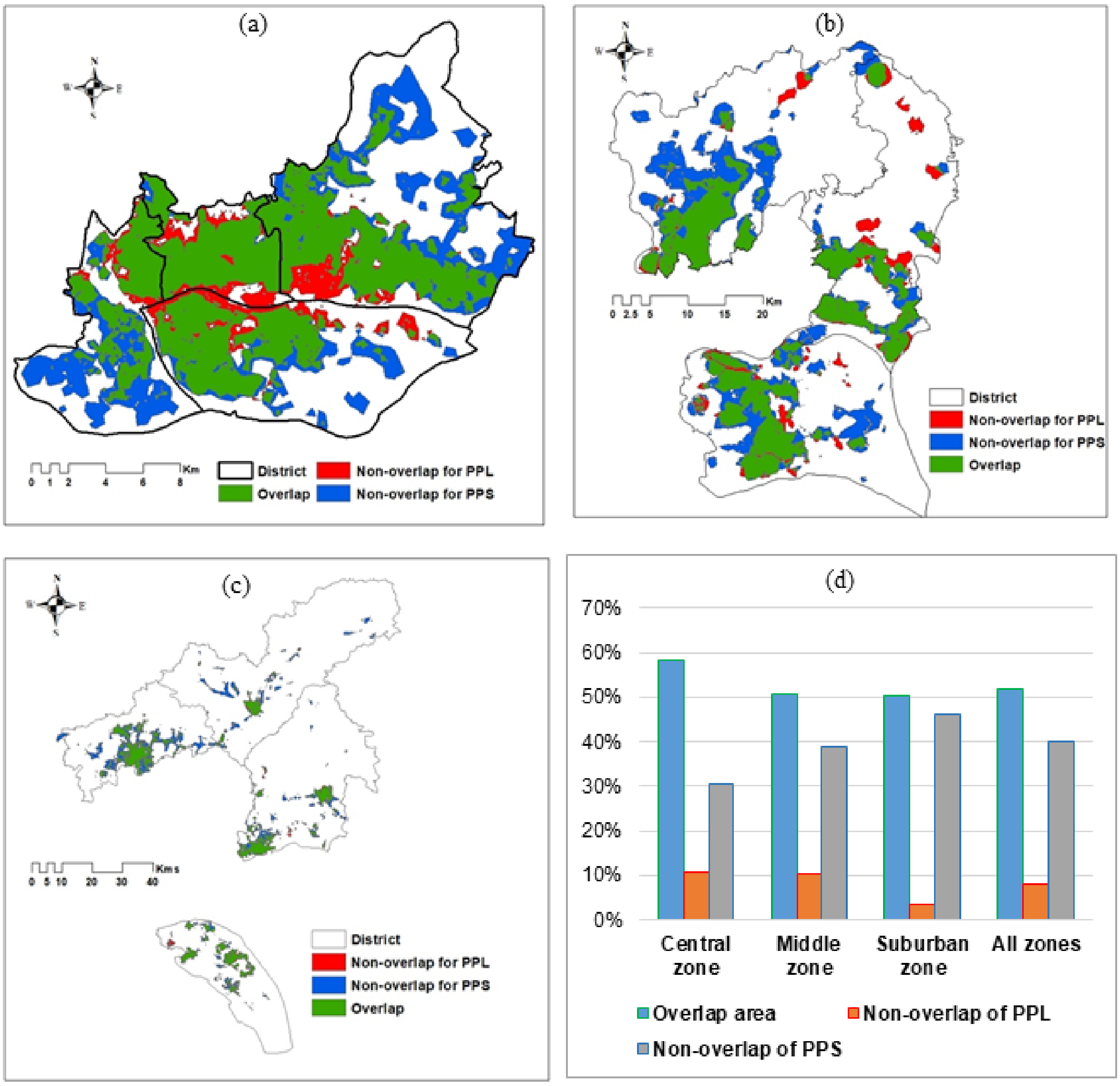



Figure 4-4. Positional relationship between the suitable areas for PPLs and PPSs in different structure zones: (a) central zone, (b) middle zone, and (c) suburban zone, and (d) Proportion of the three types of positional relationships in the suitable areas.

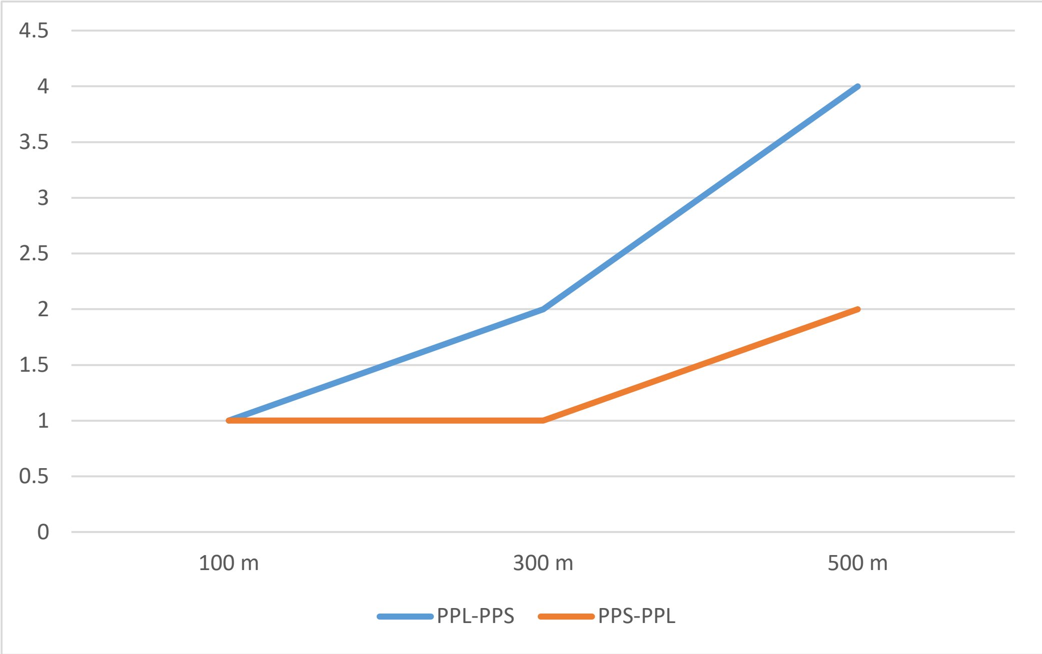
## 4.3 Coexisting relationship in the overlapping area

### 4.3.1 Compatibility with intertype competition

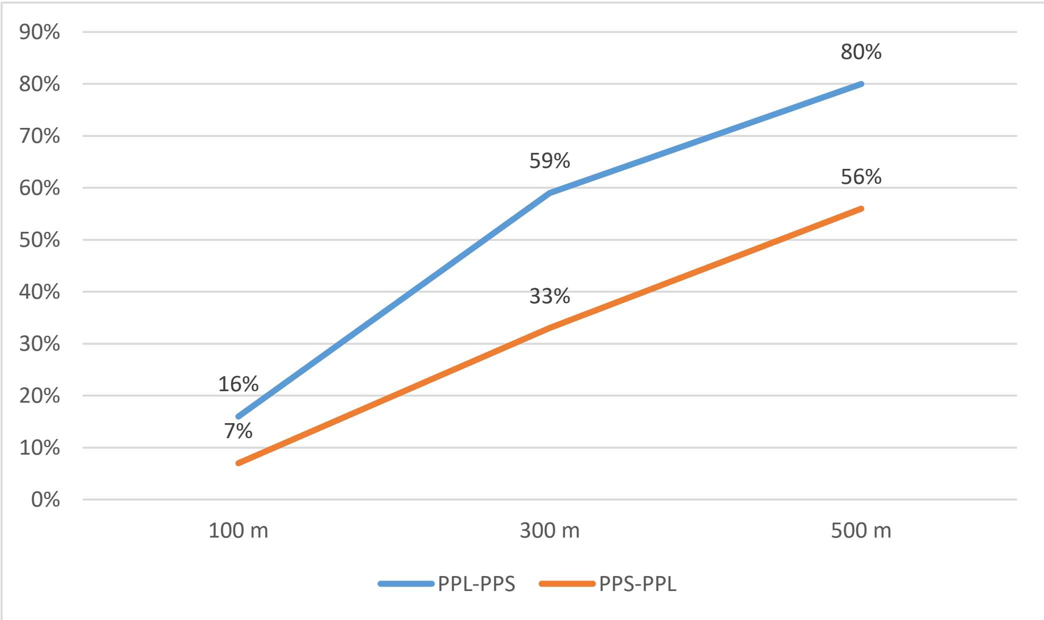
#### *4.3.1.1 Number of coexisting points*

Figure 4-5 shows the average number of coexisting points in intertype competition and the percentage of individuals with coexisting points at the three acceptable distances of the overlapping area. In the 100 m range, the average number of the coexisting PPL-PPS and PPS-PPL points is one. The percentage of coexisting PPL-PPS points is 16%, whereas that of coexisting PPS-PPL points is only 7%. The average number of coexisting PPS-PPL points remained unchanged for the 300 m distance parameter, whereas that of the coexisting PPL-PPS points increased to two. The percentage of PPL-PPS intertype competition is ~60%, whereas that of the PPS-PPL intertype competition is only 33%. In the 500 m range, the average number of coexisting PPL-PPS points is four, which is twice that of the coexisting PPS-PPL points. In this range, the PPL-PPS intertype competition is 80%. Overall, the number of coexisting points observed for the PPL-PPS intertype competition is almost twice that observed for PPS-PPL intertype competition. As the distance increased, the compatibility between the different types of PPPs increased.

As most PPLs are located in public areas inside or extremely close to the target buildings, they have more location advantages than PPSs. Customers conducting business in the target building are extremely loyal to PPLs. Consequently, PPSs should avoid competition with PPLs in a short distance. As Figure 4-5 shows, only 7% of the existing PPSs coexist with PPLs in the 100 m range, whereas 56% of existing PPSs coexist in overlap area of the 500 m range. However, 80% of existing PPLs coexist with PPSs in the 500 m range. PPL exhibits higher compatibility with intertype competition than PPS.




(a)



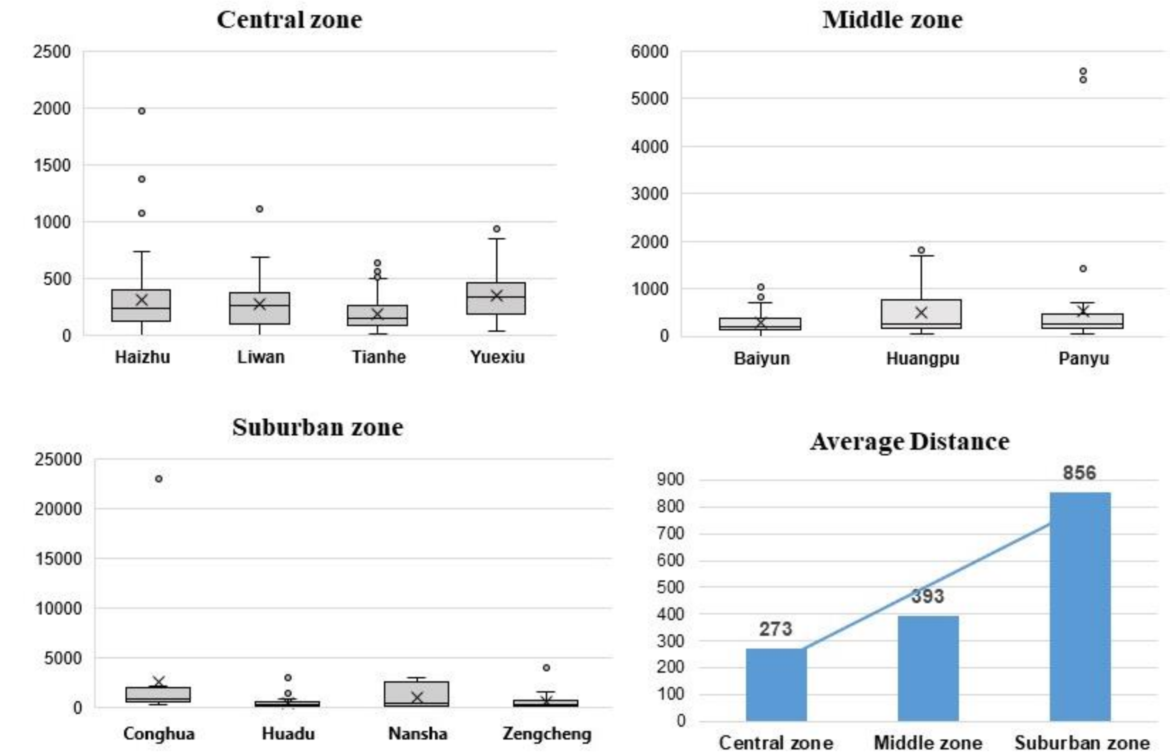

(b)

Figure 4-5. PPPs with the intertype coexisting points at the three distances of the overlapping area: (a) average number of intertype coexisting points, and (b) percentage of individuals with coexisting points.

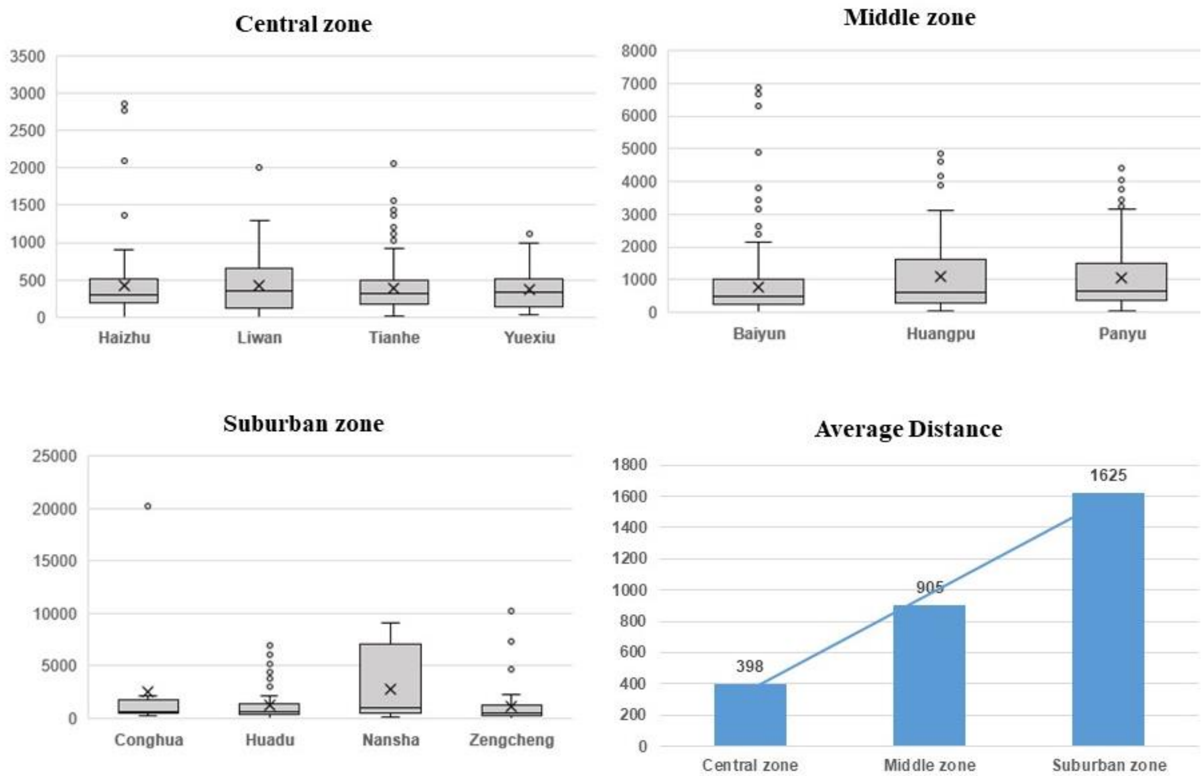
#### *4.3.1.2 Distance between the nearest coexisting points in different metropolitan structure zones*

Figure 4-6 shows the box plot of the distance of the nearest intertype points in the overlap area based on the different metropolitan structure zones. As Figure 4-6 (a) shown, the average distance between PPL and the closest PPS in the central zone is ~300 m. The distance distribution in the four administrative districts of the central zone is similar; however, Tianhe district exhibits the smallest value. In the middle zone, the average distance is ~400 m, and the Huangpu district exhibits a larger distance than the other three districts. In the suburban zone, the average distance is ~850 m, and Nansha district exhibits the largest distance. In the three metropolitan structure zones, the average distance in the central zone is the smallest; the average distances in the middle and central zones have similar values. The average distance of the suburban zone is the largest, approximately three times that of the central zone. The average distance in the three structure zones gradually increases from the central zone to the suburban zone.

As Figure 4-6 (b) shown, the average distance between PPS and the closest PPL in the three metropolitan structure zones exhibit the same trend observed for the PPL. The average distances are the smallest and largest in the central and suburban zones, respectively. However, the value of the PPS-PPL type exceeded that of the PPL-PPS type. The average distance in the central zone is approximately half that in the middle zone and one-third that in the suburban zone.




(a) PPL-PPS




(b) PPS-PPL

Figure 4-6. Box plots of the distance between the nearest points with intertype competition in the overlap area based on the different metropolitan structure zones: (a) PPS-PPL type and (b) PPL-PPS type.

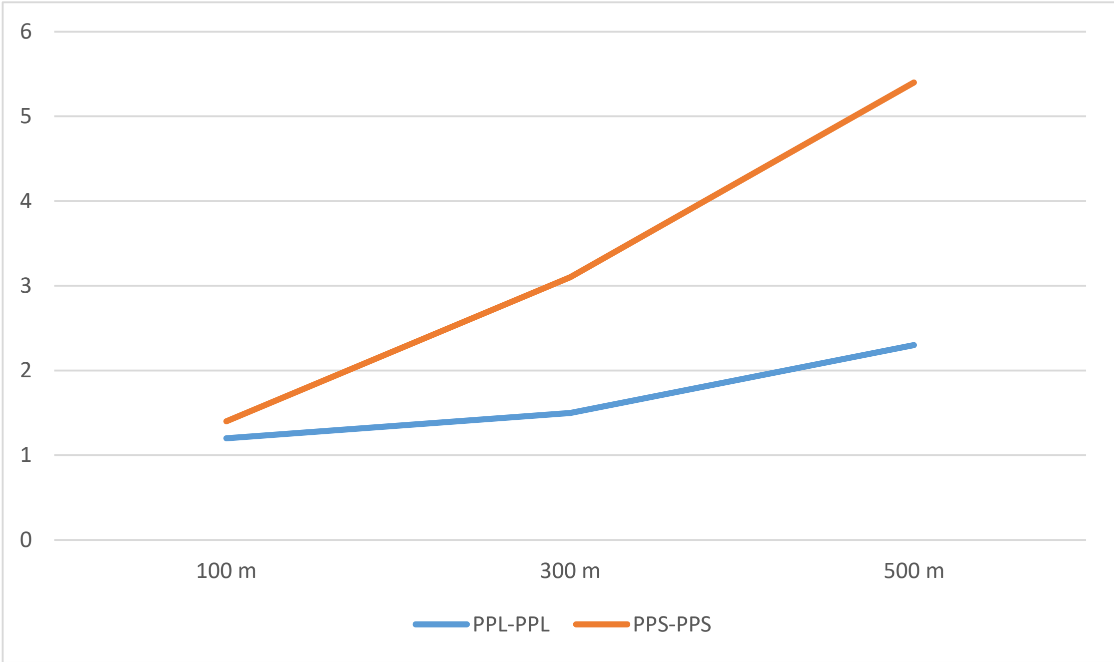
### 4.3.2 Compatibility with intratype competition

#### *4.3.2.1 Number of coexisting points*

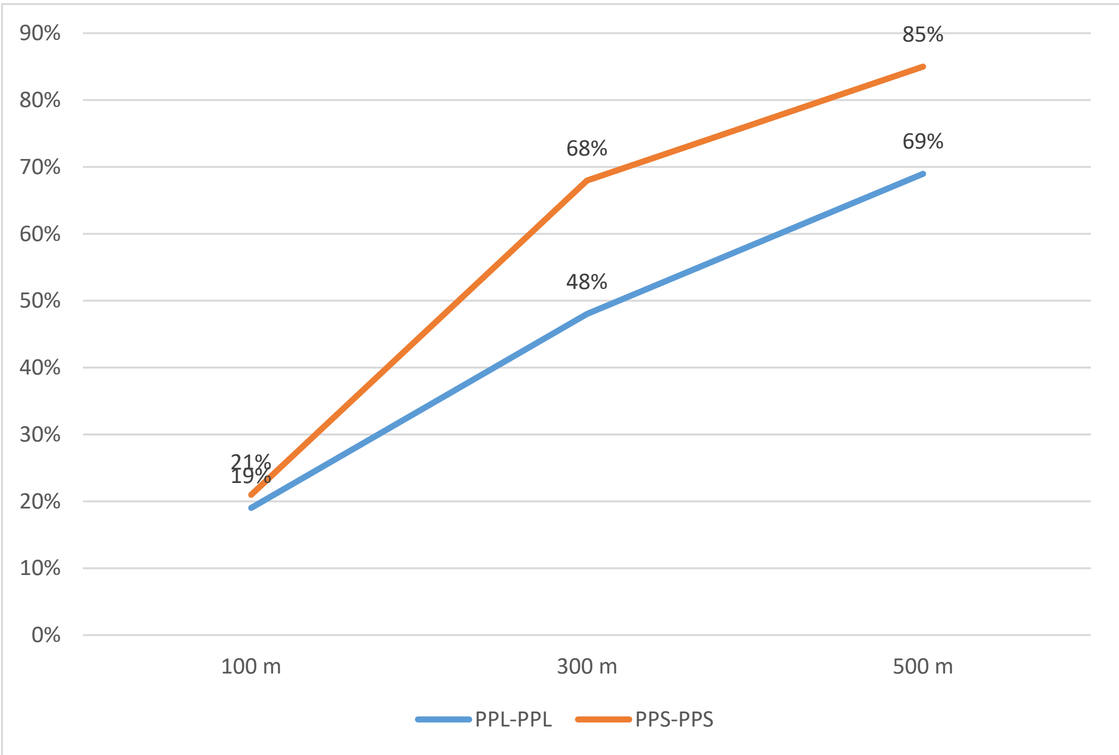
Figure 4-7 shows the average number of intratype coexisting points and the percentage of coexisting points at three acceptable distances of the overlapping area. In the 100 m range, 19% of PPL and 21% of PPS have intratype coexisting points; the average number is 1. Within 300 m, 48% of PPL have coexisting PPLs (average number: 1.5). Further, 68% of PPS have coexisting PPSs; the average number is twice that of PPL. Within 500 m, less than half of the PPL have intratype coexisting PPL points (average number: 2). Moreover, 85% of PPS have intratype coexisting points; the average number is also twice that of PPL. PPS with intratype competition exhibits a higher compatibility than PPL with intratype competition; the compatibility increases with the distance.

#### *4.3.2.2 Distance between the nearest coexisting points in different metropolitan structure zones*

Figure 4-8 depicts box plots of the distance between the nearest intratype points in the overlap area based on the different metropolitan structure zones. For the PPL, the average distance of the closest intratype PPL in the central zone is ~300 m. The average distances in the four administrative districts of this zone are similar and most are less than 500 m. In the middle zone, the average distance is 582 m. In the suburban zone, the average distance is ~1000 m. Huangpu and Nansha districts still have a large space to develop PPLs because their average distances are larger than that of the structure zone. A similar observation is noted for the intertype points. For the PPS, the average distances are ~250 m in the middle and central zones, but ~600 m in the suburban zone. The average distances of all the districts are similar to that of the structure zone.

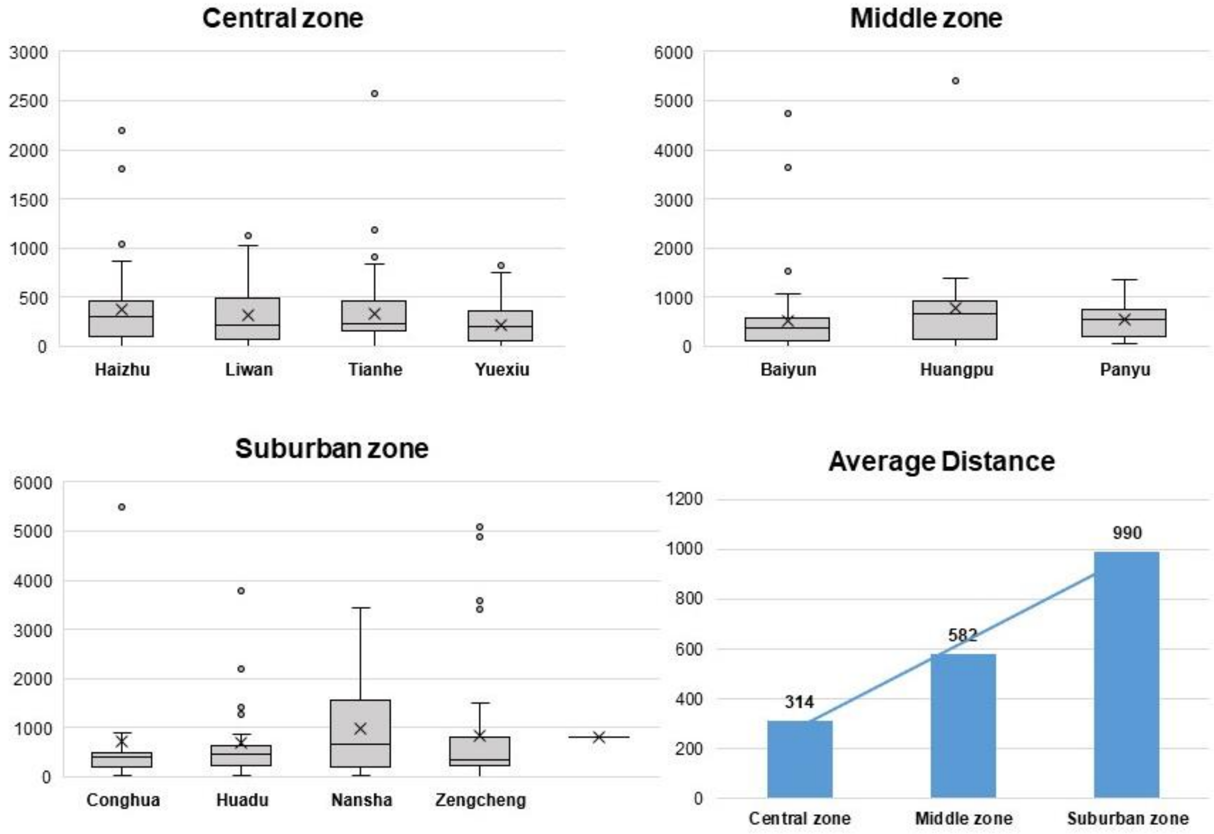



(a)

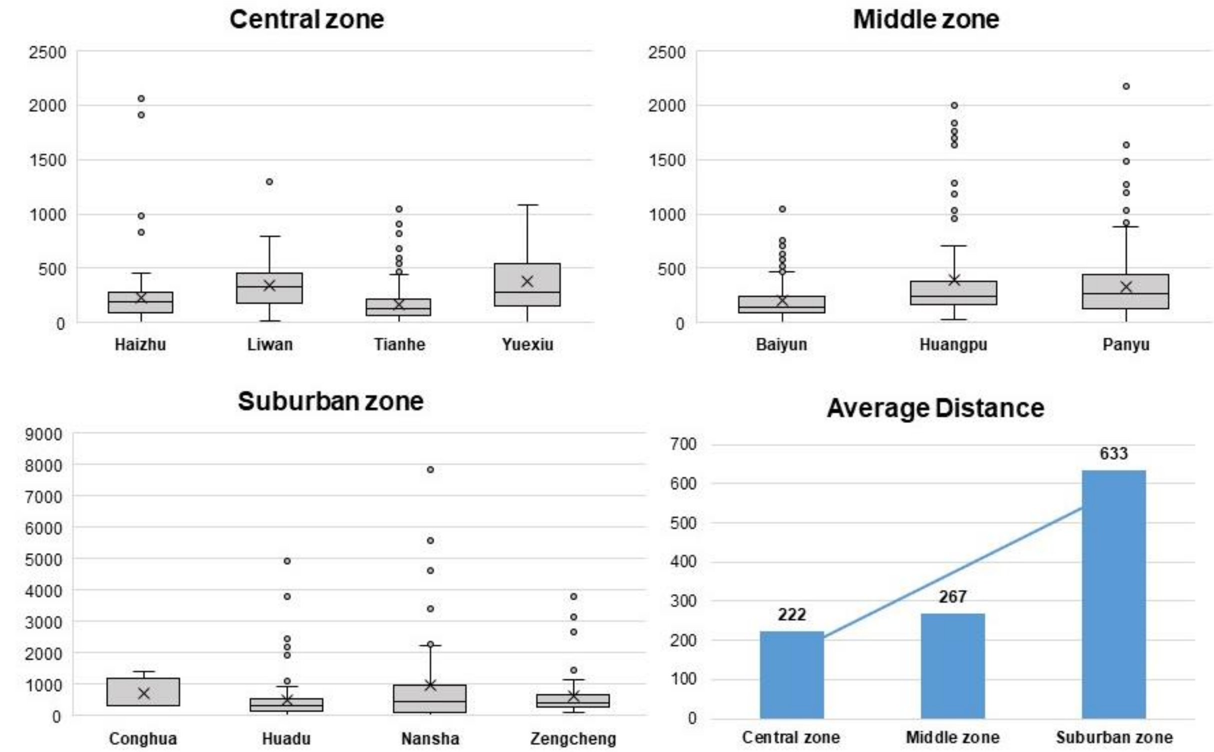



(b)

Figure 4-7. PPPs with the intratype coexisting points at the three distances of the overlap area. (a) average number of intratype coexisting points, and (b) percentage of coexisting points.




(a) PPL-PPL




(b) PPS-PPS

Figure 4-8. Box plots of the distance of the nearest intratype points in the overlap area based on different metropolitan structure zones: (a) PPL-PPL and (b) PPS-PPS.

### 4.4 Coexisting relationship in the non-overlap area

Although there was only intratype competition in the non-overlap area, resource competition will also occur when the number of PPPs reaches a certain level. Thus, in addition to analyzing the location distribution characteristics of the non-overlapping area of PPPs, it is also necessary to further explore the coexisting relationship in the intratype points to enable the study to be used as a reference to guide the planning of new points in the non-overlap area.

4.4.1 Distribution of the existing points in the non-overlap area on the multi-zone

Figure 4-9 shows the location and number of existing PPL in the non-overlap area. The total non-overlap area for PPLs in Guangzhou is 97 sq. km, with 22 existing PPL points. The total density is 0.25 points per sq. km. Of the three structure zones, the largest area of the non-overlap area is the middle zone (57 sq. km). The area with the largest number and density of existing PPLs is in the central zone. In the central zone, only the density of existing PPLs in the Liwan district is 1.2 points per sq. km, and the other three districts are all < 1 points per sq. km. In the other two zones, there are a few existing PPLs: six existing points in the middle zone and one in the suburban zone. Therefore, the symbiotic environment analysis was only conducted in the central zone of PPL. Based on the distribution of existing PPLs in the non-overlap area, additional PPLs can be developed in the non-overlap area, especially in the middle and suburban zones. The non-overlap area of PPL is only suitable for the PPL type; therefore, there is no intertype competition with PPS for resources. Hence, new PPLs should be set up first in the non-overlap areas rather than in the overlapping area.

Figure 4-10 shows the location and number of existing PPS points in the non-overlap area of the three structure zones. The total area of the non-overlap area for PPS in Guangzhou is 502 sq. km, which is approximately five times that of the non-overlap area for PPL. There are 394 existing PPSs in this area, with a density of 0.8 per sq. km. Of the three structure zones, the largest area is in the suburban zone (227 sq. km). The highest density of existing PPS is in the central zone (3 points per sq. km). The PPSs are mainly distributed in the Tianhe district. This shows that there is a small space for further PPS development. In the middle zone, the existing PPSs are primarily distributed in the Baiyun district, with a density of 1 per sq. km. The density of existing PPS in the Panyu and Huangpu districts is less than that in the middle zone. There is considerable room for the development of PPS in these two districts. In the suburban zone, the density of PPS is 0.2 points per sq. km and is mainly distributed in the Huadu district. This reveals that PPS has room for development in the non-overlap area of its niche in the middle and suburban zones with a large area and a low density of current points. In the central zone, Liwan district still has room for development owing to its relatively low density.

Overall, the area of the non-overlap area for PPS is larger than that for PPL, and the density of existing PPSs is also higher than that of existing PPLs. The new PPLs and PPSs should be preferentially arranged in their non-overlap parts of the suitable areas, particularly in the middle and suburban zones owing to the large size of the niche area and low density of existing points.

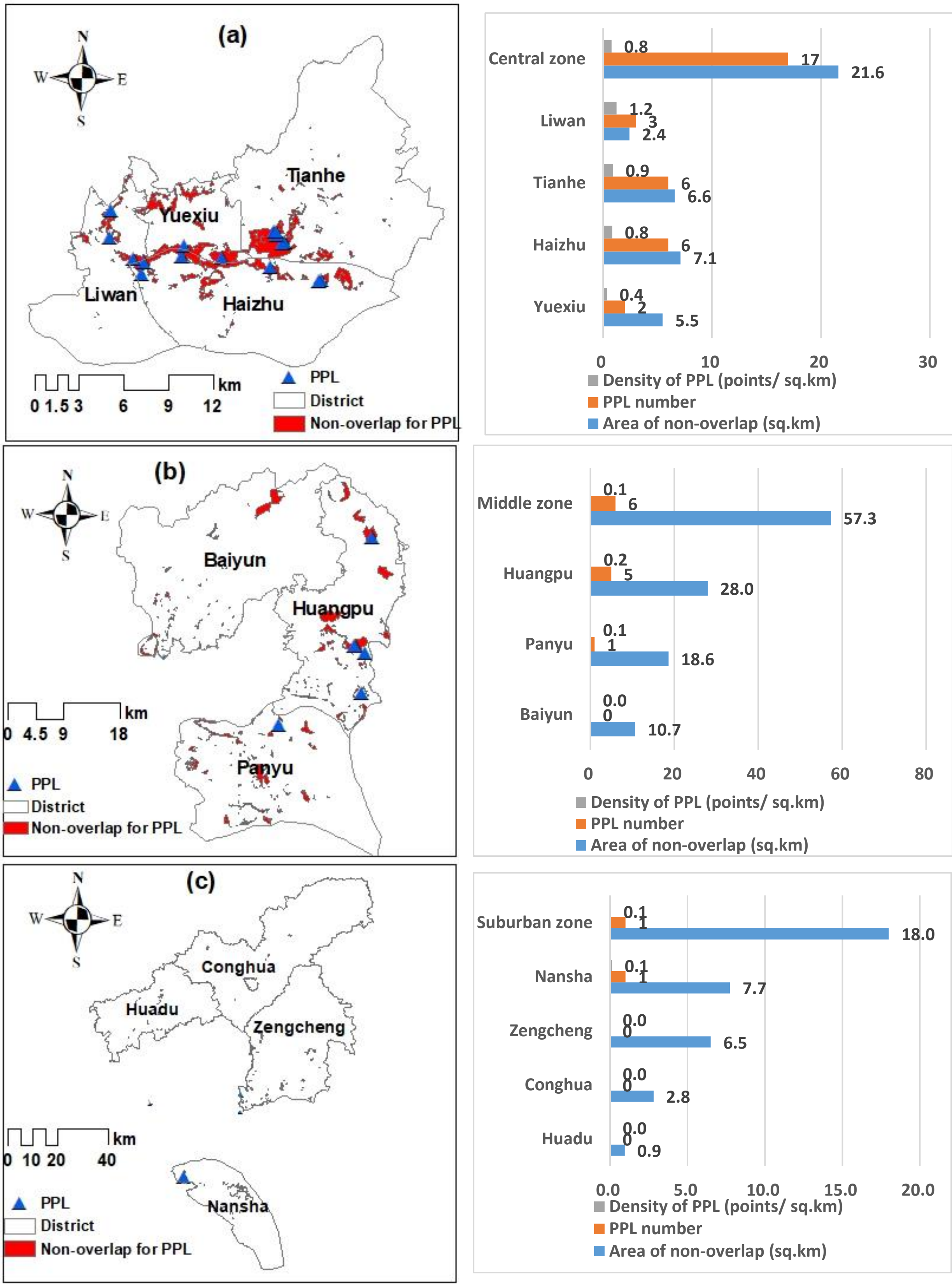



Figure 4-9. Location and number the existing PPL points in the non-overlap area of the three structure zones: (a) central zone, (b) middle zone, and (c) suburban zone.

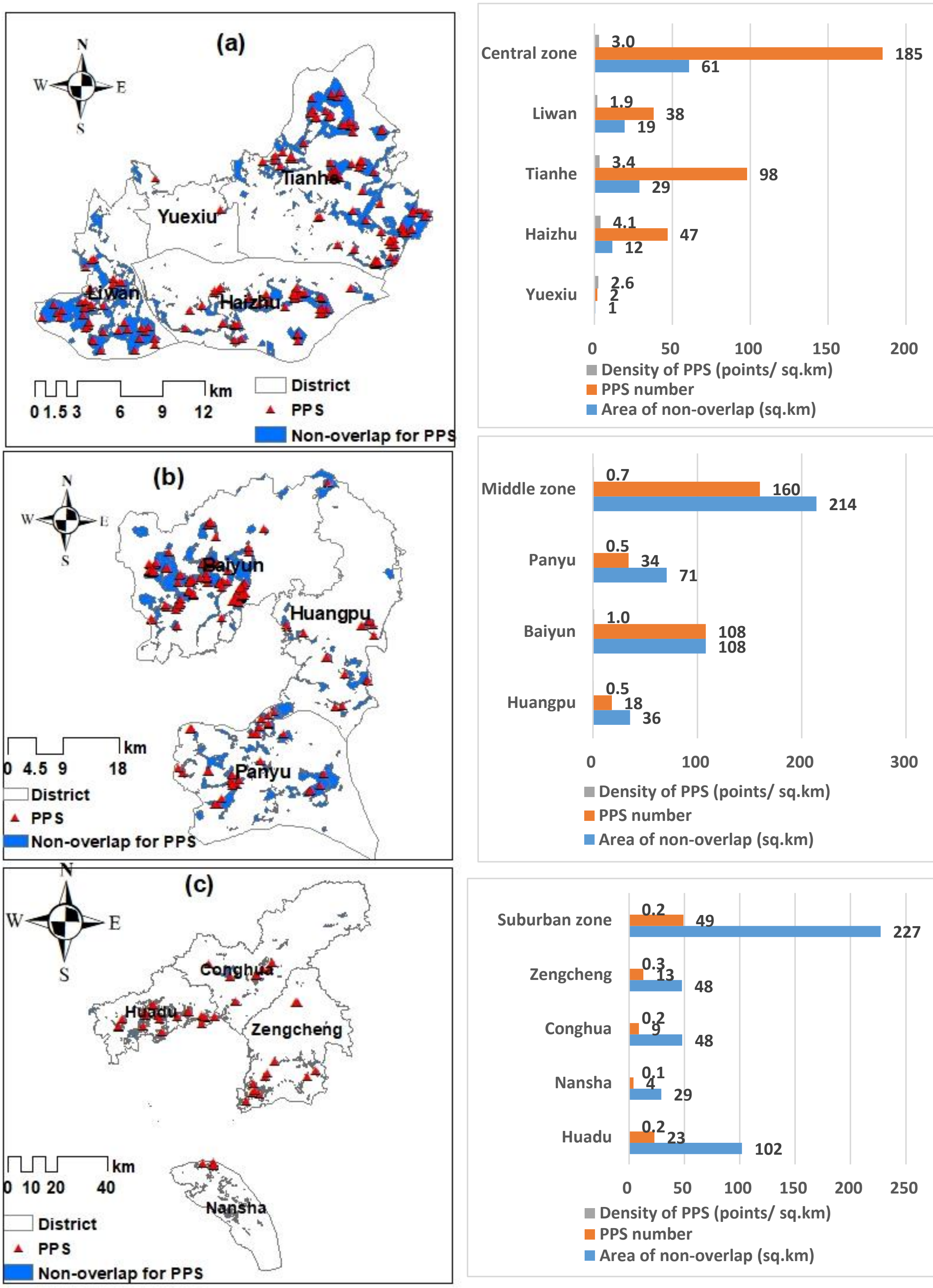



Figure 4-10. Location and number the existing PPS points in the non-overlap area of the three structure zones: (a) central zone, (b) middle zone, and (c) suburban zone.

### 4.4.2 Compatibility in the non-overlap area

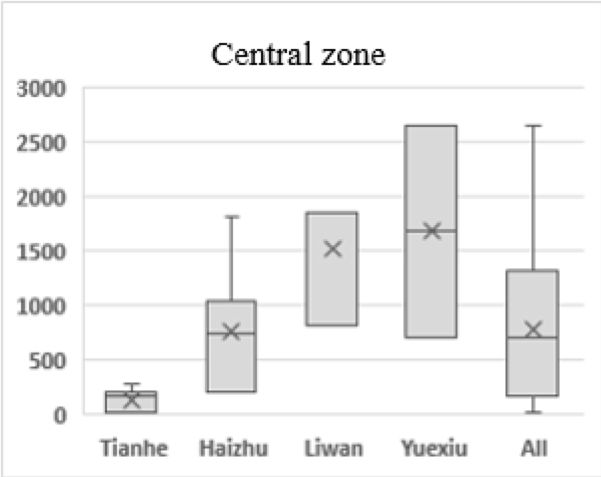
#### *4.4.2.1 Distance between the nearest coexisting points in different metropolitan structure zones*

Figure 4-11 shows the box plots of the nearest distances in the non-overlap areas of the three structure zones. For PPL, as there are few existing sites in the central and suburban zones, a detailed analysis of the symbiosis environment was not conducted. In the central zone, the average distance of the nearest PPL points is ~700 m. The shortest average distance is in Tianhe district (150 m). The largest average distance is in Yuexiu district, which is more than twice that in the central zone, followed by Liwan district. The PPL distribution in the different administrative districts in the same structural zone is not balanced. There is still considerable space for the development of PPLs in the non-overlapping areas of Yuexiu and Liwan districts in the central zone. For the PPS, the average distances to the nearest PPS in the three structure zones are approximately 400 m, 500 m, and 2000 m in the central, middle, and suburban zones, respectively. The average distance between the central and middle zones is not significantly different, but the average distance of the suburban zone is almost four times that of the central and middle zones. Therefore, as the existing PPSs are relatively far from each other in the suburban zone, there remains a large development space in the non-overlapping areas in the suburban zone. The average distance of the four administrative districts of the central area is similar to that of the central area, except for the Yuexiu district. In the middle zone, the average distances of the districts are similar. In the suburban zone, the range of the distances is large, particularly in the Conghua district.

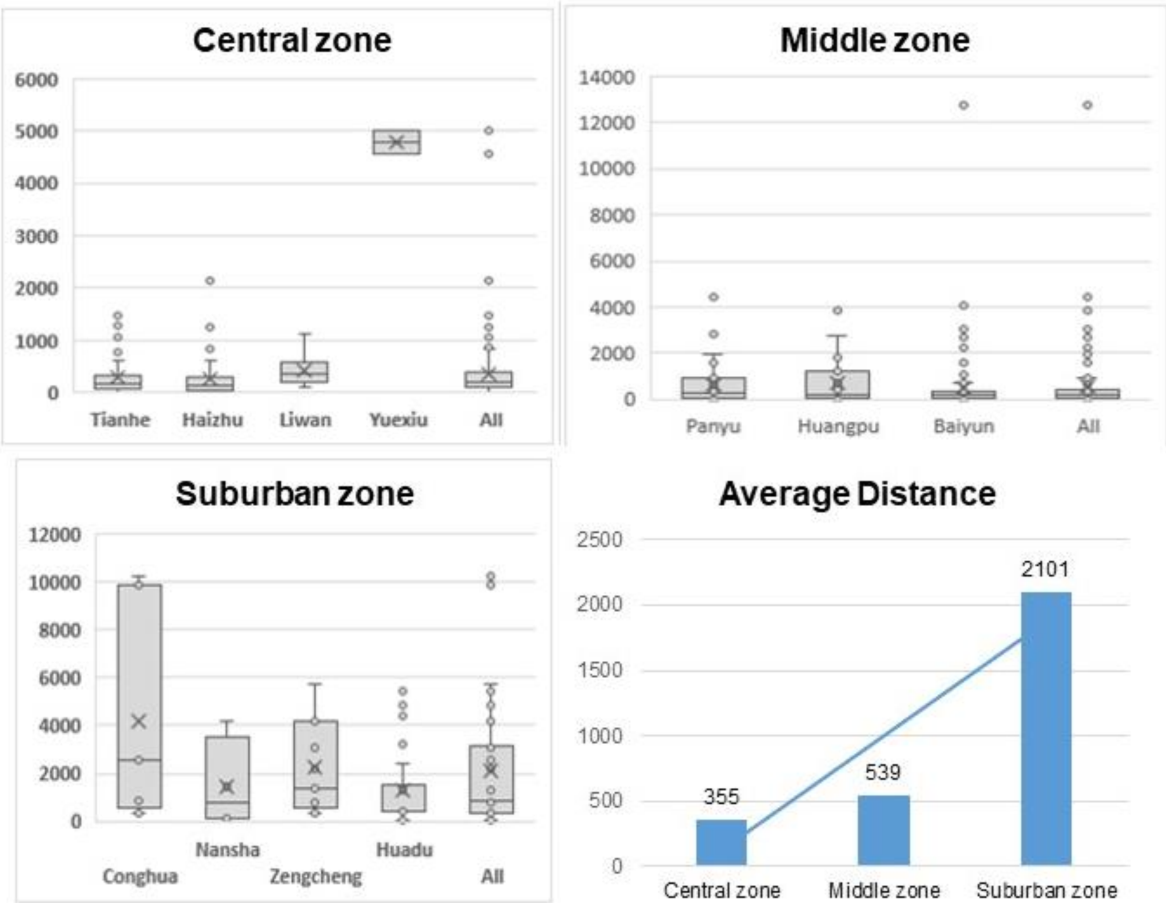
Overall, the existing PPPs are densely populated in the non-overlap area of the three structure zones; there are a few PPLs in the middle and suburban zones. Both PPLs and PPSs have considerable room for development in the suburban zone. The average distance of the nearest intratype PPL points is significantly larger than that of the intratype PPS points. For both PPLs and PPSs, the average distance is the largest in the Yuexiu district of the central zone.

#### *4.4.2.2 Number of coexisting points*

The average number of coexisting points within certain distances were analyzed to explore their coexistence. Figure 4-12 shows that the average number of coexisting PPS points in the central zone was the largest. There are approximately 4 and 2.5 points within 500 m and 300 m, respectively, corresponding to PPS and PPL in the middle and central zones, respectively. The average number of PPSs in the suburban zone decreases from 2 to 1. This demonstrates that the PPS distribution in the suburban zone is imbalanced; some points are closely distributed, but most are dispersed. At the three distances, the percentage of PPSs with coexisting points is similar in the central and the middle zones. Approximately 80%, 65%, and 30% of the coexisting PPS points are within 500 m, 300 m, and 100 m, respectively. Less than half of the points within 500 m of the suburban zone of PPS and central zone of PPL are coexisting points. This indicates that the coexisting number in the non-overlap area for PPL is considerably less than that for PPS; further, the number in the suburban zone is significantly less than those for the middle and central zones. PPSs exhibit higher compatibility in the central and middle zones. Most PPSs have coexisting PPS points within the 500 m range. Thus, the findings of this study provide quantitative data for the development of new PPPs.

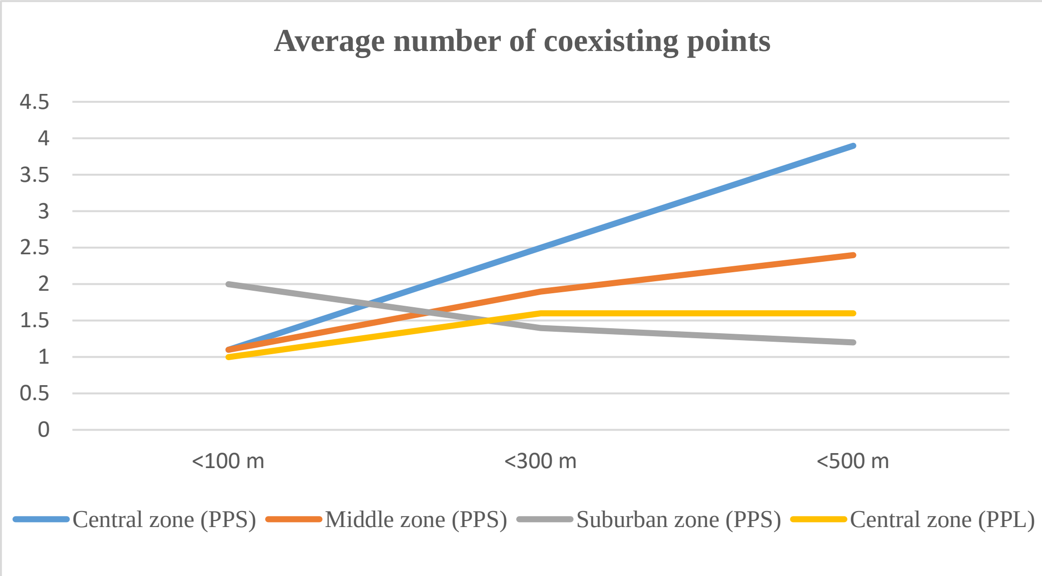



(a) PPL-PPL

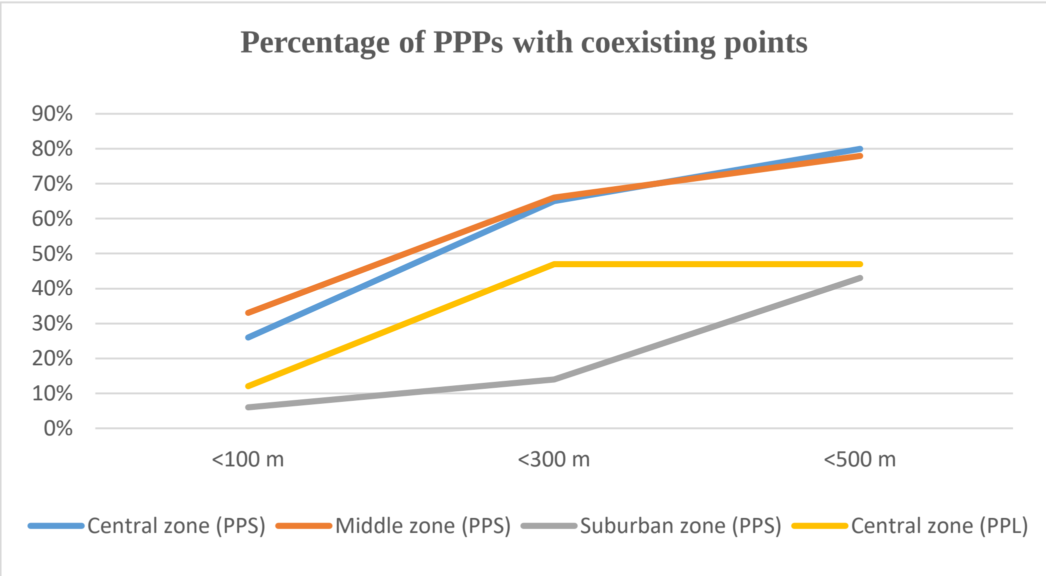



(b) PPS-PPS

Figure 4-11. Box plots of the distance between the nearest intratype points in the non-overlap area based on different metropolitan structure zones: (a) PPL-PPL and (b) PPS-PPS.




(a)



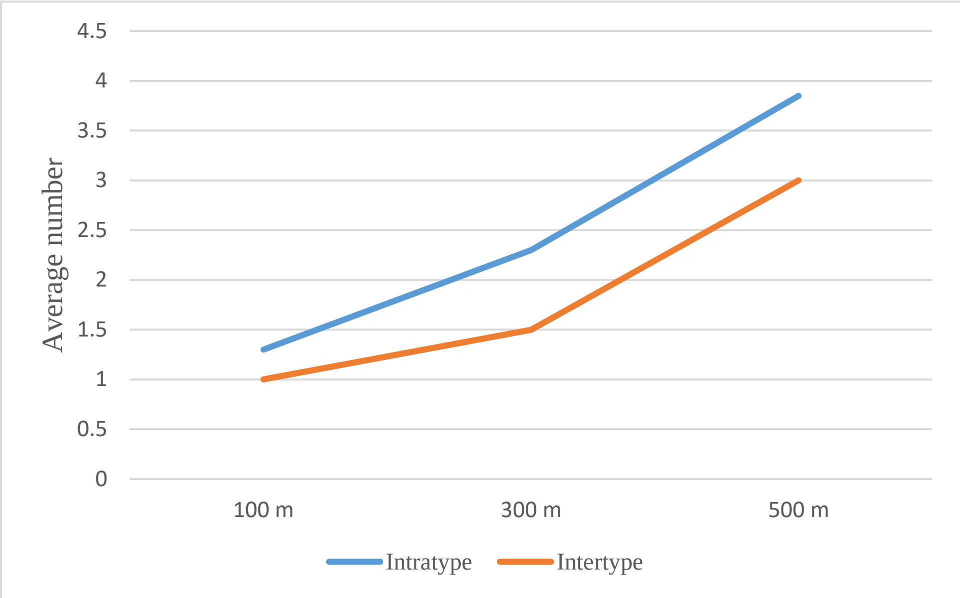

(b)

Figure 4-12. PPPs with coexisting points within certain distances in the non-overlap area: (a) average number of PPPs with coexisting points, and (b) percentage of PPPs with coexisting points.

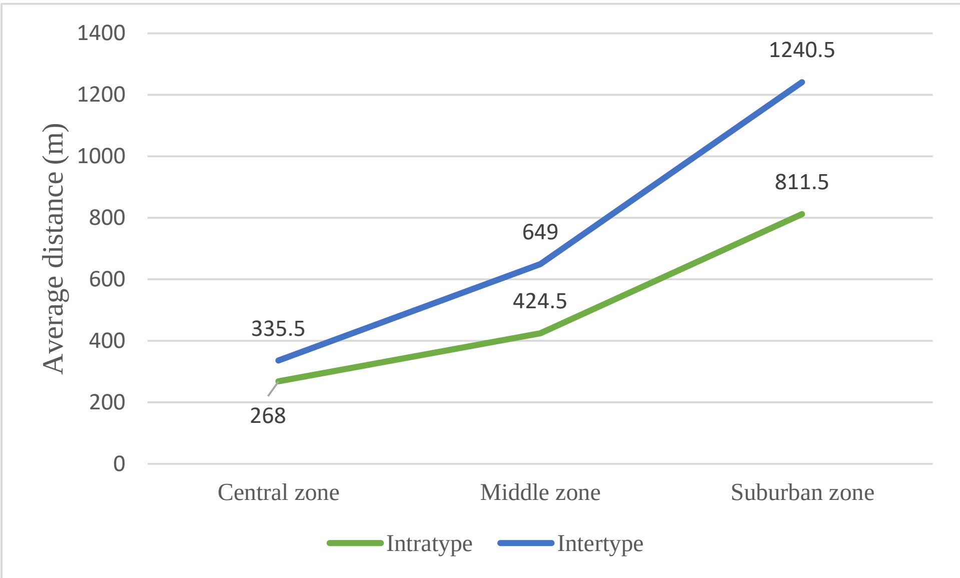
## 4.5 Discussions

To investigate the degree of compatibility with intratype and intertype competition, the average number of coexisting points and average distance to the nearest coexisting point were compared (Figure 4-13). This demonstrates that the intratype competition has more coexisting points in its surrounding and a shorter average distance to the nearest point than the intertype competition in the PPP overlapping area. This indicated that the compatibility with intratype competition was stronger than that with intertype competition, which is consistent with reality. Similar PPPs often belong to one company; therefore, the company must exercise proper management to prevent intratype competition. In addition, individuals of the same type can contribute to a scale effect, causing another type to shrink until it disappears. The coexisting distance in the intra-types of PPS is smaller than the intra-types of PPL. Intertype competition of PPL-PPS was smaller than that of PPS-PPL. Therefore, the compatibility of PPS was higher than that of PPL in the same type. The compatibility of PPL with different types of PPS was higher than that of PPS. As PPLs can be accessed 24 h a day and the security of the parcels is relatively high, PPLs are more attractive than PPSs for nearby residents. Therefore, PPSs should be preferentially located in regions far from PPLs.

PPSs are more compatible with intratype competition than PPLs. Conversely, PPLs are more compatible with intertype competition than PPSs. The main reason for this observation is that the number of existing PPSs is considerably more than that of existing PPLs. Consequently, a large number of the same type of PPS are located near each other and fewer PPLs are located farther apart. There are more PPSs than PPLs in the PPL surroundings within a shorter distance. In reality, PPLs have more limitations than PPSs. First, PPL is a machine facility that does not utilize human interaction; therefore, the elderly may experience challenges in operating. Second, the cabinet has a limited capacity and uniform size, which may not be sufficient for the volume and size of various parcels. Third, the cost and maintenance fee of PPL facilities are extremely high. PPS is relatively more flexible than PPL. Even in the overlap areas where two types of competition exist, many PPSs are located around PPLs and coexist peacefully. The compatibility of PPL for interspecies is higher than that for PPS. However, PPLs should be positioned within a reasonable number and distance between two PPP types.




(a)




(b)

Figure 4-13. Comparing the compatibility of intertype and intratype competition in the overlapping area: (a) average number of coexisting points, and (b) average distance to the nearest points.

## 4.6 Summary

In chapter 4, the niche overlap theory in ecology was applied to analyze the spatial relationship of the suitable areas and coexistence between the two PPP types. The suitability areas for the two PPP types were simulated in Chapter 3. The spatial relationship of the suitable areas for PPL and PPS partially overlapped. The suitable areas can be separated into three parts— overlap area, non-overlap area of PPL, and non-overlap area of PPS. The overlap area is considerably larger than the non-overlap of the two PPPs types, particularly in the central zone. The non-overlap of PPS is significantly larger than that of PPL, particularly in the suburban zone.

According to the niche overlap theory, when two niches overlap, it will be difficult for the two species to coexist for a long time owing to the competition and exclusion, unless space and resources are abundant. As the layout of existing PPPs results from several years of market competition verification, it is assumed that the existing PPPs coexist. This chapter analyzes the coexistence of the two types of PPPs based on two aspects: the number of coexisting PPPs within a certain distance and the nearest distance of coexisting PPPs. The high compatibility indicates the presence of more coexisting PPPs in the surrounding and a short distance to the nearest coexisting PPP. In the overlapping area, the PPP competition includes both intertype and intratype competition. In the non-overlapping area, intratype competition was more prominent.

The study confirmed that, in the overlapping area, the intertype competition of PPL-PPS was more compatible than that of PPS-PPL, whereas the intratype competition of PPS was more compatible than that of PPL. In the two types of competition, intratype competition was more compatible than intertype competition. In the non-overlap area, PPS was more compatible than PPL. Of the three metropolitan structure zones, the compatibility was highest and lowest in the central and suburban zones, respectively. The compatibility of the metropolitan structure zones was related to urbanization.

The interaction of the two PPP types should be considered to select the sites for PPPs. Priority is given to selecting the non-overlapping areas to avoid intertype competition. Locations with a large area without existing PPPs or a low density of existing points with a large degree of urbanization should be selected because the central zone has the highest compatibility in the three structure zones. Appropriate spacing should be selected based on the number and distance of the coexisting species to position the PPPs in suitable areas to avoid competition. As the intertype competition of PPL-PPS has a higher compatibility than that of PPS-PPL and the intratype competition is more compatible than intertype competition, the new PPSs should be arranged around existing PPSs, avoiding areas near PPLs.

# Chapter 5 Layout strategy of PPPs

## 5.1 Differences between PPS and PPL locations

The multi-zone LR model described in Chapter 3 selected the crucial factors from 27 candidate variables to simulate PPP suitability. The results revealed that different factors were selected in the different structure zones. Figure 5-1 summarizes the similarities and differences between the location factors of the two PPP types based on the metropolitan structure zones. The factors shown in the two ellipses are the important for the two PPP types. The overlapping areas of the ellipses show the factors selected in both PPP types. The factors in blue and black represent positive and negative correlation factors, respectively. The location differences between the two types of PPPs were analyzed based on six characteristics.

### 5.1.1 Differences in the impact of land price

The SLPrice factor had no effect on the three structure zones of PPL. The PPS type selected this factor in both the central and middle zones. This was related to the manner in which the two types of PPPs were developed. PPLs are mainly located in the public area of residential and commercial areas. They are typically established as a supporting facility. The rent of the land occupied by the facility is not in line with the market price. Instead, it is usually leased at a lower rate than the standard land price. The owners of the buildings aim to improve the supporting facilities and increase customer satisfaction by establishing nonprofit PPLs to rent. The cost of the PPL mainly comprises the initial cost of the facility and the operating cost, with the site fee only accounting for a small portion. This is in keeping with the actual situation that the standard land price has little effect on PPL site selection. PPSs mainly use the parasitic business model in retail stores. The rent of the store is the largest expenditure of commercial operations. Therefore, land price is important for PPSs, and is typically higher in the central and middle zones. Shops with high land prices usually have good locations with a high people flow. These shops are typically unwilling to share expensive land to increase the flow of people and income. Land prices are relatively high in the metropolitan central and middle zones, but relatively low in the suburban zone. Therefore, the model for the PPS suitable area only selected land price as a factor in the first two structural areas. Land price was a negative correlation factor; therefore, the lower the land price, the higher the suitability.

### 5.1.2 Differences in the main service objects

In the model for the suitable area of the two types of PPPs, the co-selected building factor was dist_Resi_Qua. Thus, residential quarters were the service targets of the two types of PPPs. A residential quarter is also called a residential community or neighborhood. It refers to a largely residential area with a relatively independent living environment in a certain area of the city and is equipped with a complete set of living service facilities (Zhang *et al*., 2019). It is characterized by a relatively closed and independent area, high population density, complete supporting facilities, and a property management organization. The residents mostly have property rights of their house. The PPLs are developed by negotiating with the property management organization to build on the community land, while PPSs cooperate with the shops supporting the community. In addition to residential quarters, another service object of the PPS is dormitories. This factor was selected in the three structural zones using the multi-zone LR model. Dormitory residential areas mainly include school dormitories and dormitories in factories or units. Dormitories have higher population mobility and are relatively more difficult to manage than residential quarters. There is usually no particular property management company, and the supporting facilities are incomplete. PPL did not select the commercial office buildings factor instead of the dormitory factor in the three structural zones. This type of building is usually well managed by the property management company.

In general, the main service objects of PPL are residential communities and commercial office buildings. Conversely, the main PPS locations are residential quarters and dormitories.

### 5.1.3 Differences in the impact of population density

As shown in Figure 5-1, PPL and PPS both selected the population density factor in the suburban zone, and PPS also selected it in the middle zone. The importance of the population factor was almost the lowest of the selected factors in both PPL and PPS, as shown in Tables 3-9 and 3-10. This indicates that the population factor is not very important in the PPP suitability simulation model. The population density data source used was the population prediction data in the WorldPop dataset developed by the WorldPop Project (https://www.worldpop.org). The dataset provides annual gridded population data from the 2000–2020 period, with a spatial resolution of 100 m. The input variables of WorldPop included the most recent official census population data and nighttime satellite images using a random forest regression tree-based mapping approach (Gaughan *et al.*, 2016).

In the literature review, the previous reports proposed that population density had a positive correlation with PPPs but did not mention that the daytime and nighttime populations had different impacts on the two types of PPPs. Moreover, the population density in most studies used the population density of the administrative district from the government census data, which is also the nighttime population data. As daytime population data are difficult to obtain in China, this study aims to use the commercial building location corresponding to the daytime population to replace the daytime population analysis, which can reduce the difficulty of data collection. The accuracy of the suitability simulation revealed the feasibility of this method. Meanwhile, the importance of population density is also related to the scale of research. This study uses a micro-scale pixel as the research unit, which is different from previous studies using administrative divisions as the unit. This demonstrated that the importance of population density is lower when a smaller unit is used as a research unit.

### 5.1.4 Differences in the influence of road factors

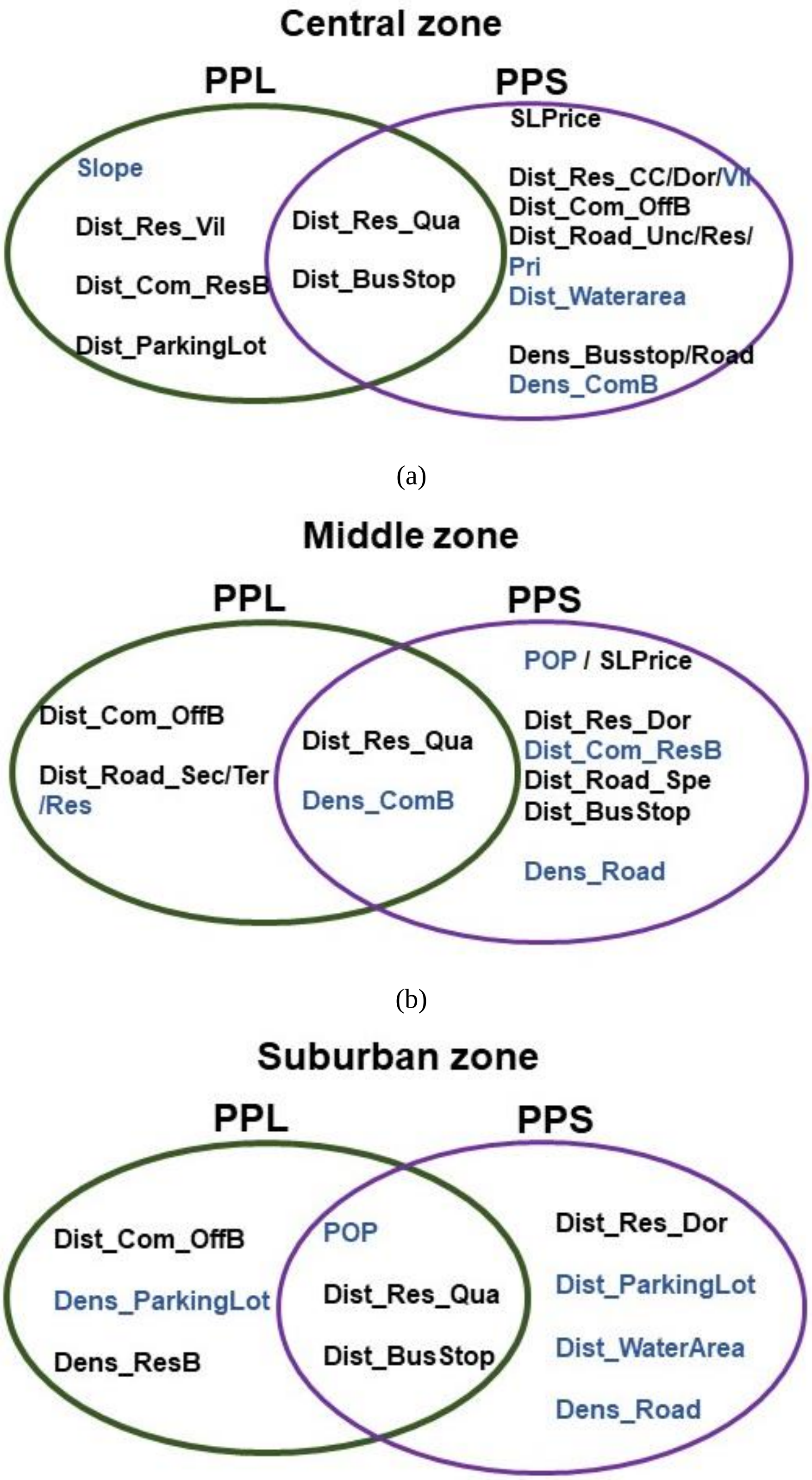
The road factor is refined into seven road types, from the primary roads to paths, to explore the various road factors related to the PPP location. Using the standard LR model of the suitability simulation, the PPL type only selected the nearest distance factor of the secondary road, whereas the PPS type selected the closest distance factor of residential roads, special roads, and paths. This revealed that the roads selected by PPS all had few lanes, narrow road width, and were suitable for walking. The roads selected by PPL were the main roads of the city and were suitable for vehicles. Using the multi-zone LR model, the road distance factors were not selected by PPLs in the central and suburban zones. In the middle zone, the secondary, tertiary, and residential type roads were selected. Only the residential road distance exhibited a positive correlation; the remainder exhibited a negative correlation. Thus, the farther away from the residential road and the closer to the secondary and tertiary roads, the more suitable it is. In the PPS type, the road distance factors were not selected in the suburban zone. The unclassed and residential road types were selected as the negatively correlated factors and the primary road type was selected as the positively correlated factor in the central zone. The special road type was selected as the negatively correlated factor in the middle zone. Overall, the road distance factors were more critical in the middle zone in both PPL and PPS, possibly owing to the characteristics of the middle zone. Secondary and tertiary roads were critical for PPL, while small roads close to residential areas were critical for the PPS layout. The road distance factors had no effect on the two types of PPPs in the suburban zone. Further, these factors had a weaker effect on PPL than on PPS.

5.1.5 Differences in the influence of transportation

The PPP location model did not select the metro station factor in any zones owing to the relatively small number of subway stations and the extremely high land prices nearby. The bus station was selected as the means of transportation with the widest coverage and largest population. In addition, the nearest parking lot distance factor was also selected in the central zone of the PPL location model, and its importance is higher than the nearest bus station distance. The three structural areas for PPS selected the nearest bus station distance, and the importance reduced gradually from the suburbs to the middle area to the downtown area. Bus stations with a large flow of people are also suitable locations for PPPs. Furthermore, near to the parking lots are also suitable locations for PPLs in the central zone. This may be related to the way in which customers collect parcels.

5.1.6 Differences in facility property

The main service objects of PPS are residential quarters and dormitory buildings. PPSs can be developed near residential roads, paths, other walkable roads, or near bus stations with low land prices. PPSs are open to all customers and are extremely accessible. The main service objects of PPL are residential quarters and commercial office buildings. PPLs can be developed near surrounding second and tertiary roads or bus stations and parking lots. PPLs are typically established by signing an agreement with the manager or developer of the target buildings as a supporting facility. This is a win-win situation. PPLs can obtain lower rents, and it is convenient to the customer. PPLs are usually located inside the private area of the target buildings, with low openness. PPLs are more inclined to public service facilities, whereas PPSs are more inclined to commercial service facilities. The nature of the facility determines its profit model. PPLs have been cultivating customer behavior and have provided free services to the public since their establishment. If they suddenly convert to commercial profitable models and charge users, customer dissatisfaction and resistance will result. Moreover, this would lead the customers to utilize PPSs. When the supply-demand relationship changes, PPSs will continue to increase the service area by setting more points, which will widen the gap between PPLs with high investment costs and a small number. This may even cause PPLs to go to extinct. This is similar to the previously shared bicycle service. Many funds were invested in the early stages to purchase equipment and cultivate the habit of launching free or low price strategies. Later, this model could not withstand the long-term losses and high operating costs. Once the capital chain is broken, the PPLs will fail. The Chinese government began to realize this when online shopping continued to increase sharply as the pandemic progressed, and PPLs, which allowed the contactless collection of goods, played a vital role in preventing the spread of COVID-19. In April 2020, the inclusion of PPLs in public service facilities to support land use and subsidy policies was proposed.




(a)



(b)

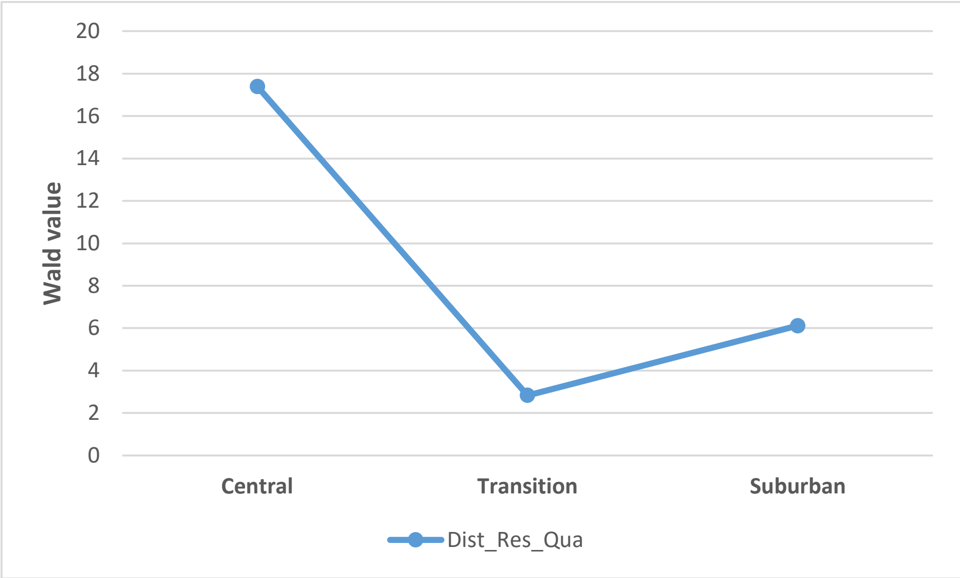


(c)

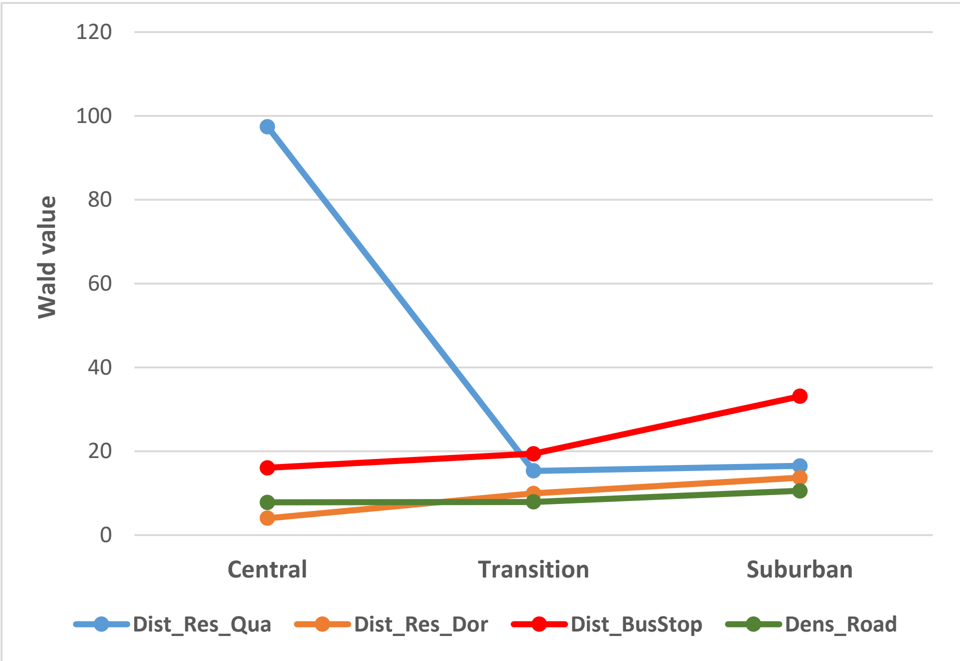
Figure 5-1. Differences between the location factors of the two PPP types in the metropolitan structure zones: (a) central zone, (b) middle zone, and (c) suburban zone.

## 5.2 Impact of metropolitan structure zone on PPP location

In the previous section, the differences between PPS and PPL locations were analyzed. The same type of PPPs in the different metropolitan structures have different location characteristics, which is reflected in the different factors selected by the multi-zones model and their different importance. Furthermore, the importance differed, even when the same impact factors were selected. Four factors were selected for PPSs in all the metropolitan structure zones: Dist_Res_Qua, Dist_Res_Dor, Dist_BusStop, and Dens_Road. Figure 6-2 shows that the importance of Dist_Res_Qua gradually decreased from the city center to the suburbs, while the remaining Dist_Res_Dor, Dist_BusStop, and Dens_Road gradually increased. In particular, Dist_BusStop has become the most critical factor in the suburbs. Only the Dist_Res_Qua factor was selected for PPLs in the three metropolitan structure areas; its importance also decreased from the central area to the suburbs. The most critical factors in the middle zone and the suburbs are Dist_Road_Ter and Dist_BusStop. Bus stops and roads are a part of the city infrastructure. In the city center, the infrastructure is full, and most of the houses appear in the form of residential quarters. There are more farmer houses in the middle zone and suburban zone and fewer residential high-rise clusters. PPPs were established near infrastructures to satisfy the wide demands. Additionally, many people work in the central area and live in the middle or suburban zones. The convenience of roads and public transportation are predominant conditions for population increase. As shown in Figure 5-3, the PPP location in the three metropolitan structure zones is characterized by the fact that the most critical factor in the central zone is close to the target service building factor, and the most critical factor in the middle zone and suburban zone is close to the infrastructure.




(a) PPL




(b) PPS

Figure 5-2. Importance of the common factors of PPPs in the three metropolitan structure zones: (a) PPL and (b) PPS.

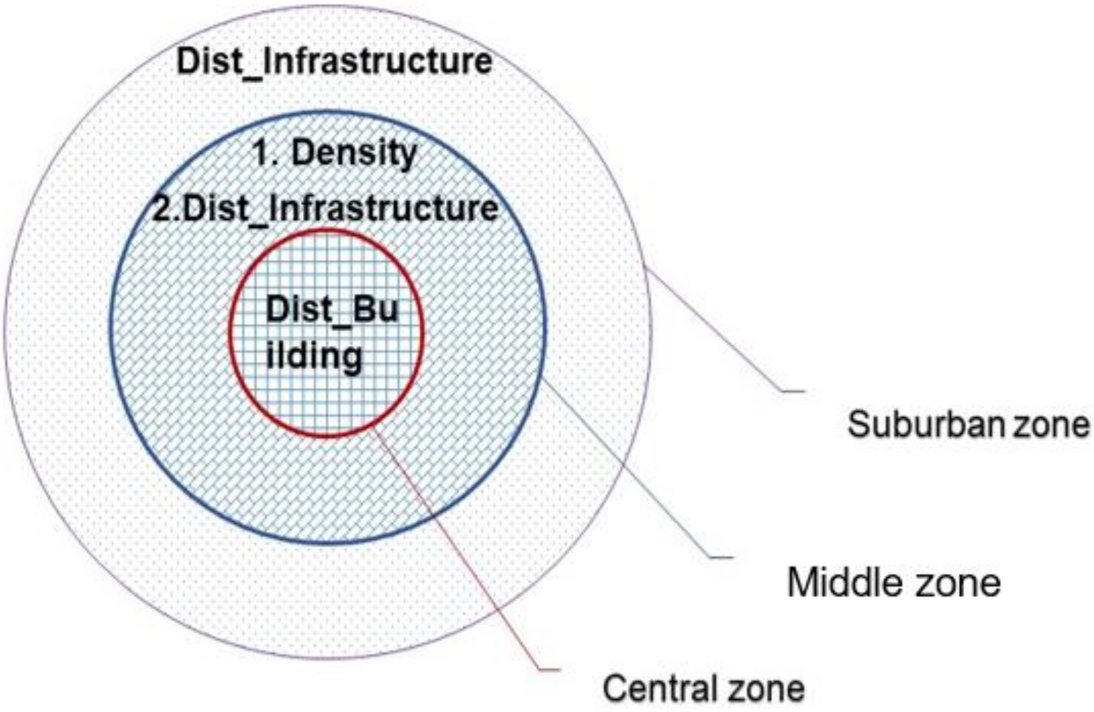



Figure 5-3. The most important factors for PPP locations in different metropolitan structure zones.

### 5.3 Layout strategy considering the mutual compatibility of the two PPP types

Based on the results of Chapter 4, PPL shows high compatibility with PPS, but PPS has low compatibility with PPL. PPL can be considered as a supplement to PPS. Further, PPSs are based on a shop-in-shop concept because it cooperates with existing stores or self-built stores by adding delivery services. If there is no physical store location, PPSs cannot be established. PPLs have no physical restrictions and can be established in any open space in the city. Therefore, the planning of PPS facilities is prioritized. PPLs can be planned in places where PPSs cannot be established or where the existing PPS facilities do not meet the demand. The distance between the facilities should not be too close, and should be appropriately spaced to prevent competition.

The relationship of suitable areas for the two types of PPPs was partially overlapped. Intratype and intertype competition existed in the overlapping area. In the non-overlapping area, there was only intratype competition. Prioritizing site selection in non-overlapping areas can prevent intratype competition. The development of PPL in the non-overlapping area was relatively large owing to the low density of the existing points. PPS had higher intratype compatibility in the non-overlapping areas due to more coexisting PPS and a shorter distance to the nearest coexisting PPPs. There was still considerable room to identify cooperative stores to develop new PPSs. The central zone had the highest compatibility in the metropolitan structure zones, followed by the middle and the suburban zones. In the central zone of the PPS non-overlapping area, the number of existing PPS facilities in Yuexiu District was relatively low, and the distance was significantly smaller than the average distance of the central zone. Thus, this region was a suitable area for PPSs. Figure 6-4 shows an example of the layout strategy of PPS in the non-overlap area, which is in the northwest of Yuexiu District without existing PPS facilities. This area mainly comprised two parts. On the upper left is the Huifu International Business Center, where a large number of office buildings for clothing, leather shoes, decoration materials, and other related industries gathered. The eastern area is the place where the Guangzhou railway is repaired and maintained, and below is the staff dormitory of the railway department. This is entirely consistent with the location characteristics of the main service objects of PPS in the central zone, including business office buildings and dormitories, analyzed previously. It is also one of the positional differences between PPS and PPL. When planning PPP locations for overlapping areas, PPPs should be preferentially developed in the central and middle areas because of their more excellent compatibility. Of the various types of competition in the overlap area, PPS has the highest compatibility with intratype competition, followed by PPL-PPS intertype competition. The lowest compatibility was with the PPS-PPL intertype competition. Therefore, when selecting PPS sites in the overlapping area, PPSs can be developed around existing PPSs, but are not suitable near PPLs. When selecting the PPL location in the overlapping area, PPLs can be developed around PPSs but should be a certain distance away from existing PPLs.

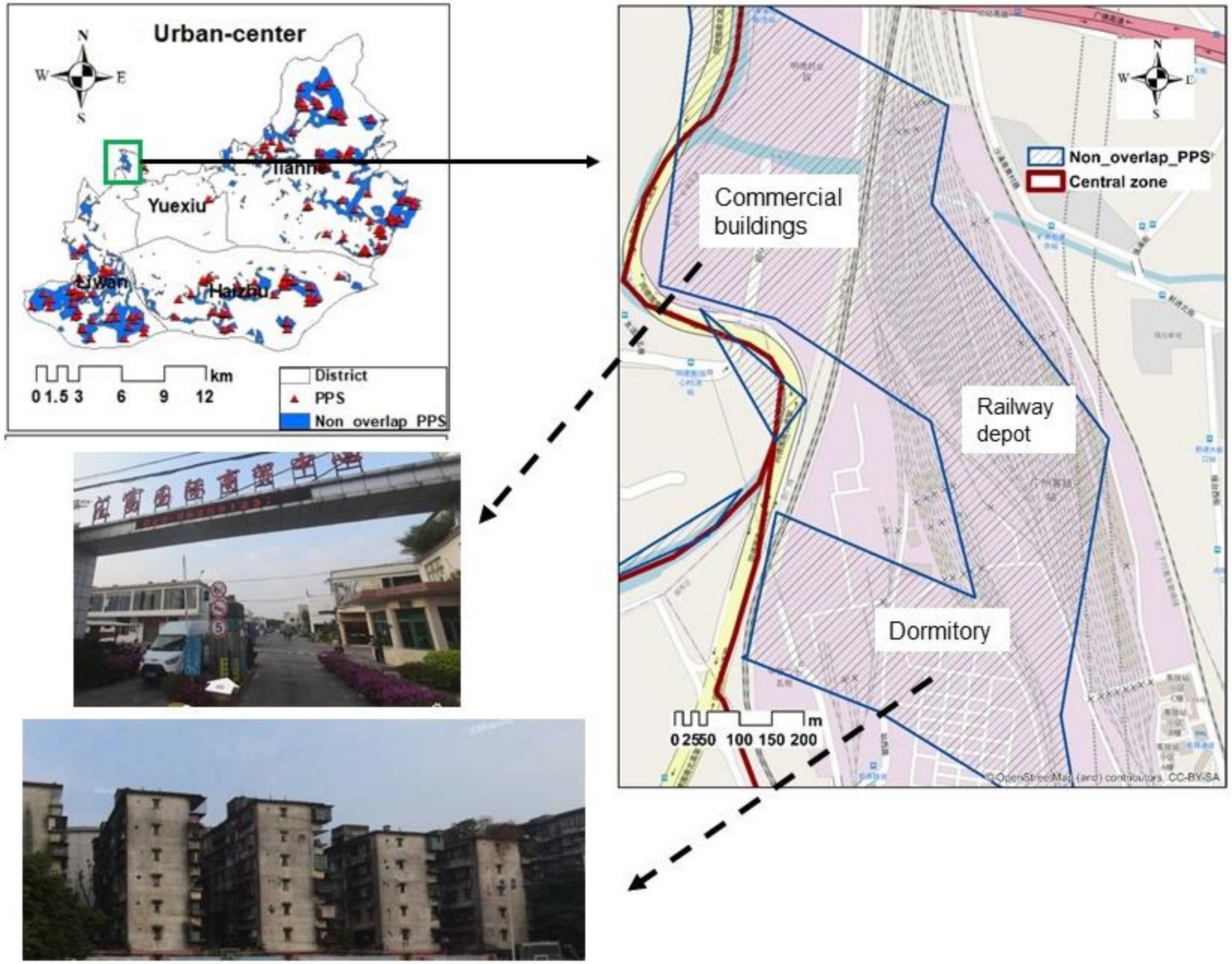



Figure 5-4. Example of the layout strategy of PPS in the non-overlap area.

# Chapter 6 Conclusions

The rapid development of e-commerce has severely impacted logistics distribution, and the last-mile delivery problem has restricted logistics development. PPPs are the most effective, widely used novel solution that helps firms to reduce costs through consolidated shipments and provide customers with a flexible, convenient, and comfortable means of receiving parcels. PPPs have significant research interest in logistics research and offer numerous advantages to different fields (Gevaers *et al*., 2014; Jung *et al*., 2006; Kämäräinen *et al*., 2001; Maere, 2017; Taniguchi and Kakimoto, 2003). There are limited studies on the facility arrangement from a geographic perspective. Previous reports state that PPP location is strongly related to the population and spatial accessibility (Li *et al*., 2018; Li *et al*., 2019; Lin *et al*., 2019; Morganti *et al*., 2014a; Tan *et al*., 2016; Xue *et al*., 2019). However, these reports qualitatively analyzed the factors affecting the layout of PPPs and rarely used quantitative methods to explore the PPP location and layout. Further, there are two types of PPPs, but most studies only analyzed one and ignored the interaction between them. There is currently no guideline for decision-makers to determine the suitable areas for the two types of PPP separately and suitably arrange them to prevent competition. Furthermore, in quantitative studies, the study area is mainly in a small residential area, and there are limited studies on the metropolitan area scale. Thus, this study aims to identify the location differences between the two types of PPPs in the Guangzhou metropolitan area and the layout strategy considering the interaction between the two types. The study analyzed the spatial distribution characteristic with the administrative division unit and simulated the suitable location area of PPPs with the grid unit. Based on the models selected, the location differences between the two types of PPPs were analyzed. The coexisting relationship between the two types of PPPs and the same type of PPPs was analyzed in the simulated suitable location area. The study provides a guide and reference for decision-makers to rationally plan the two types of PPP to avoid mutual competition and achieve sustainability.

There are six chapters in this study. Population density is an important index of metropolitan area structure division (Dickinson, 2013) and is also an important factor for PPP location (Morganti *et al*., 2014a). According to the characteristics of Guangzhou metropolitan area development and population density, the 11 administrative divisions are divided into three metropolitan structure zones: central, middle, and suburban. Chapter 2 described different spatial quantitative methods to analyze the spatial agglomeration pattern and the spatial correlation for PPPs. The distribution of PPPs was imbalanced in the Guangzhou metropolitan area. The PPP distributions in the three metropolitan structural zones were extremely different. The densest area was the central zone, followed by the middle zone. Few PPPs were dispersed in the suburban zone. This is similar to the pattern of the population of Guangzhou. It reveals that the distribution characteristic of PPPs is the density, which decreased from the central zone to the suburban zone. The degree of SA also gradually decreased from the suburban zone to the central zone in both PPL and PPS. In the suburban zone, the main spatial association type was the LL cluster. In the middle zone, the percentage of the LL cluster decreased. In the central zone, the HH cluster and LH outlier type were staggered, especially the PPS. This demonstrated that there was apparent spatial heterogeneity of the PPS distribution in the central zone. Moreover, the distribution characteristic of the two types of PPPs was also significantly different. In the administrative districts of all the structure zones, the greater the population density of the district, the greater the PPL density. The district in the central zone for PPS did not follow this trend. The PPS density was abnormal in areas with large differences in population density between day and night. Owing to the different distribution characteristics of the two types of PPPs in the three metropolitan structural zones, Chapter 3 described the simulation of the two types of PPPs with a standard and a multi-zone model and selection of the better performance model to generate suitability maps for PPPs. Chapter 3 explored the relationship between the two types of PPP and surrounding geographic factors based on the simulation of the suitability of PPPs using the LR model of ML. Pixels were used as a research unit and a PPP reference database was built with the 27 variables using big data. The data was divided into the training and test datasets. The LR model was used to learn the data characteristics from the training dataset and determine the best parameters of the classification model. The test dataset checked the accuracy of the model. The results found that the multi-zone LR model performed better than the standard LR model. The suitability maps for the two types of PPPs in the Guangzhou metropolitan area were generated using the best model. The results showed that the suitable areas for PPS were larger than that for PPL. The most important variable affecting the distribution of the two types of PPPs in the standard

LR model was Dist_Res_Qua. In chapter 4, the niche overlap theory in ecology was applied to analyze the spatial relationship of the suitable areas generated from Chapter 3 and the coexisting relationship of the two PPP types. The spatial relationship of the suitable areas for PPL and PPS was partially overlapping. The suitable areas can be separated into three parts: overlapping area, non-overlap area of PPL, and non-overlap area of PPS. The overlap area was much larger than the non-overlap of the two PPP types, especially in the central zone. The non-overlap of PPS was much larger than that of PPL, especially in the suburban zone. The study found that in the overlapping area, the intertype competition of PPL-PPS had a higher compatibility than that of PPS-PPL, and the intratype competition of PPS had higher compatibility than that of PPL. Of the two types of competition, the compatibility with the intratype was higher than that with the intratype. In the non-overlap area, the compatibility of PPS was higher than that of PPL. Of the three metropolitan structure zones, the compatibility of the central zone was the highest and that of the suburban zone was the lowest. The compatibility of the metropolitan structure zones was related to its urbanization. Chapter 5 analyzed the location difference of the two types of PPPs, the impact of metropolitan structure zone on the PPP location, and the layout strategy considering the mutual compatibility of the two PPP types.

There are three novel findings of this study. First, the ecological niche overlap theory is applied to the location analysis for the sustainable development of PPPs. To avoid competition between PPPs, the two types of PPPs should be used as a whole system for location planning considering their interaction. The coexisting relationship between the intertype and intratype competition was analyzed in the three metropolitan structure zones. It provides a data reference of the appropriate spacing for the two types of PPPs in the suitable areas. Second, the study conducted a suitability simulation of PPPs in a metropolitan area using the LR model of ML and compared the performance of the standard and multi-zone LR models. Third, this study used big data to identify the specific factors for PPP locations. The population-related factors were refined to the four residential building types and two commercial types. The accessibility factor was refined to seven road types and three transportation nodes (bus stop, metro exit, and parking lot).

This study has several contributions. First, the location differences of two types of PPPs were analyzed based on six characteristics—main service objects, facility attributes, impact of land price, road factors, transportation, and population. The location differences between the two types of PPPs are that PPLs are close to commercial buildings in the middle and suburban zones, while PPSs are close to commercial buildings in the central zone. PPS is close to the dormitory residential buildings in the three structural zones. In previous studies, it was merely proposed that population has a positive correlation with PPPs, but did not mention that the daytime and nighttime populations have a different impact on the two types of PPPs. Besides, because the daytime population data are difficult to obtain, this study used commercial building locations corresponding to the daytime population to replace the daytime population analysis, which can reduce the difficulty and increase the accuracy. Land price does not affect the three structure zones of PPL but affects PPS in the central and middle zones. It is related to the way in which the two types of facilities were established. In terms of transportation, PPL and PPS are close to the bus stops. Also, the parking lots are a suitable location for PPLs in the central zone. In terms of roads, secondary and tertiary type roads are critical for PPL, while small roads close to residential areas are critical for PPS layout. The pick-up method for PPSs is mainly by foot, so it is suitable to arrange them in a convenient walking area. The location characteristics and establishment methods of PPL and PPS revealed that PPLs are similar to public service facilities, while PPSs are similar to commercial service facilities. Second, the study elucidated the interaction of two types of PPPs. In the overlapping area, the intertype competition of PPL-PPS had a higher compatibility than that of PPS-PPL, and the intratype competition of PPS had a higher compatibility than that of PPL. Of the two types of competition, the compatibility with intratype was higher than with intertype. In the non-overlap area, the compatibility of PPS was higher than that of PPL. In the three metropolitan structure zones, the compatibility was highest in the central zone and lowest in the suburban zone. The compatibility of the metropolitan structure zones is related to its urbanization. PPL can be considered as a supplement to PPS. Third, the LR model of ML performed well in the suitability simulation of PPPs, and the multi-zone LR model was better than the standard LR model. A metropolitan area is a region consisting of a densely populated urban core and its less-populated surrounding territories, and population was the important factor for PPPs. Multiple structure zones were preferred for the simulation and the result was more accurate when the study area was macro-scale with an imbalanced population density. Fourth, the result of the suitable simulation model indicated that the suitable area of PPSs accounts for 16.5% of the total area of Guangzhou, while PPL accounts for 10.7%. The allocation of PPPs can be focused on these suitable areas. It significantly reduced the analysis difficulty and time spent by decision-makers. Fifth, the structure zones have an impact on the PPPs location. PPP location in the three metropolitan structure zones is characterized by the fact that the most critical factor in the central zone is close to the service target building factor, and the most critical factor in the middle zone and the suburban zone is close to the infrastructure. When selecting locations for PPPs in metropolitan areas, the characteristics of the metropolitan structure zones should also be considered and conducted according to the planning strategy. This was not proposed in previous studies.

Despite the contributions of the research, there are still some limitations. First, the metropolitan structure zones classification is based on the population density and the development of the metropolitan area. Although population density is one of the important indicators for the division of urban structure, many cultural and economic factors also need to be analyzed. The division of urban structure will be further investigated in future research. Second, the suitability simulation model did not consider the willingness and demand of customers. Third, the results are obtained only from the analysis of one typical metropolitan area. Whether the same can be applied in other metropolitan areas requires further verification. In future research, the PPPs in the multiple metropolitan areas will be analyzed to determine the similarities and differences. The calculated demand population will be used instead of the total population variable to improve the accuracy of the model.

# Acknowledgements

I would like to express my gratitude to all those who have given me assistance, help and support during the completion of my Ph.D. in Division of Spatial and Information Science(SIS), Graduate School of Life and Environment Sciences, University of Tsukuba, Japan.

Foremost, I wish to express my deepest and sincere gratitude to my academic supervisor, Assistant Professor Takehiro MORIMOTO, for his constant guidance, invaluable inspiration, encouragement, and continuous support. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D. study. I would like to express my heartfelt thanks to Professor Yuji MURAYAMA, Professor Keisuke MATSUI, and Professor Jun TSUTSUMI, members of the dissertation advisory committee for their constructive suggestions that immensely helped me to improve this research.

Finally, I am extremely grateful to my daughter, parents, and my family members for their continuous encouragement and moral support I received throughout my studies. It gave me the strength to achieve my goal and overcome various difficulties.

# References


Akıncı, H., Özalp, A. Y., & Turgut, B. (2013). Agricultural land use suitability analysis using GIS and AHP technique. *Computers and Electronics in Agriculture,* 97, 71-82.

Anselin, L. (2003). An introduction to spatial autocorrelation analysis with GeoDa. http://geodacenter.asu.edu/system/files/spauto_0.pdf (accessed 10.03.2021).

Balcik, B., Beamon, B. M., & Smilowitz, K. (2008). Last mile distribution in humanitarian relief. *Journal of Intelligent Transportation Systems*, *12*(2), 51-63.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: John Wiley & Sons.

Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: the Bonferroni method. *Bmj*, *310*(6973), 170.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1-3.

Burgess E.W. (1924). The growth of the city: an introduction to a research project. *Publications of the American Sociological Society*, *18,* 85-97.

Cárdenas, I., Beckers, J., & Vanelslander, T. (2017). E-commerce last-mile in Belgium: Developing an external cost delivery index. *Research in Transportation Business & Management*, *24*, 123-129.

Celemín, J. P., & Velázquez, G. Á. (2018). Spatial analysis of the relationship between a life quality index, HDI and poverty in the province of Buenos Aires and the autonomous city of Buenos Aires, Argentina. *Social Indicators Research*, *140*(1), 57-77.

Cui, C., Wang, J., Wu, Z., Ni, J., & Qian, T. (2016). The socio-spatial distribution of leisure venues: A case study of karaoke bars in Nanjing, China. *ISPRS International Journal of Geo-Information*, *5*(9), 150.

Dickinson, R. E. (2013). *City, Region and Regionalism: A geographical contribution to human ecology*. London: Routledge and Kegan Paul Ltd.

Edwards, J. B., McKinnon, A. C., & Cullinane, S. L. (2010). Comparative analysis of the carbon footprints of conventional and online retailing: A. *International Journal of Physical Distribution & Logistics Management*, *40*(1-2), 103-123.

Flack, V. F., & Chang, P. C. (1987). Frequency of selecting noise variables in subset regression analysis: a simulation study. *The American Statistician*, *41*(1), 84-86.

Gallo, J.L., & Ertur, C. (2003). Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980–1995. *Papers in Regional Science*, *82*(2), 175-201.

Gaughan, A.E., Stevens, F.R., Huang, Z., Nieves, J.J., Sorichetta, A., Lai, S., Ye, X., Linard, C., Hornby, G.M., Hay, S.I., Yu, H., & Tatem, A.J. (2016). Spatiotemporal patterns of population in mainland China, 1990 to 2010. *Scientific Data*, *3*(1), 1-11.

Gevaers, R., Van de Voorde, E., & Vanelslander, T. (2009). Characteristics of innovations in last-mile logistics-using best practices, case studies and making the link with green and sustainable logistics. *Association for European Transport and Contributors*, 1-21.

Gevaers, R., Van de Voorde, E., & Vanelslander, T. (2014). Cost modelling and simulation of last-mile characteristics in an innovative B2C supply chain environment with implications on metropolitan areas and cities. *Procedia-Social and Behavioral Sciences*, *125*, 398-411.

Giller, P. S. (1984). Niche theory. In *Community Structure and the Niche* (pp. 8-21). Dordrecht: Springer.

Huang,T. (2017). *Study on the Layout of Express Self-pickup Network Based on GIS*. Xi’an: Chang’an University. (in Chinese)

Hopkins, L. D. (1977). Methods for generating land suitability maps: a comparative evaluation. *Journal of the American Institute of Planners*, *43*(4), 386-400.

Hosmer, D.W., & Lemeshow S.S. (1989). *Applied Logistic Regression*. New York: John Wiley & Sons.

Iwan, S., Kijewska, K., & Lemke, J. (2016). Analysis of parcel lockers' efficiency as the last mile delivery solution–the results of the research in Poland. *Transportation Research Procedia, 12*, 644-655.

Jung, H., Lee, K., & Chun, W. (2006). Integration of GIS, GPS, and optimization technologies for the effective control of parcel delivery service. *Computers & Industrial Engineering*, *51*(1), 154-162.

Kämäräinen, V., Saranen, J., & Holmström, J. (2001). The reception box impact on home delivery efficiency in the e-grocery business. *International Journal of Physical Distribution & Logistics Management, 31*(6), 414-426.

Kroll, C. N., & Song, P. (2013). Impact of multicollinearity on small sample hydrologic regression models. *Water Resources Research*, *49*(6), 3756-3769.

Lachapelle, U., Burke, M., Brotherton, A., & Leung, A. (2018). Parcel locker systems in a car dominant city: Location, characterisation and potential impacts on city planning and consumer travel access. *Journal of Transport Geography*, *71*, 1-14.

Lee, K.-I., & Koval, J. J. (1997). Determination of the best significance level in forward stepwise logistic regression. *Communications in Statistics-Simulation and Computation*, *26*(2), 559-575.

Li, G., Chen, W. & Yang, L. (2019). Spatial pattern and agglomeration mode of parcel collection and delivery points in Wuhan City. *Progress in Geography, 38*(3), 407-416. (in Chinese)

Li, G., Yang, L., & He, J. (2018). The spatial pattern and organization relation of the pickup points based on POI data in Xi'an: focus on Cainiao stations. *Scientia Geographica Sinica*, *38*(12), 2024–2030. (in Chinese)

Li, N. (2013). *Research on the Layout of the Self Pick up Points of KB Company*. Beijing: Beijing Jiaotong University. (in Chinese)

Lin, L., Han, H., Yan, W., Nakayama, S., & Shu, X. (2019). Measuring Spatial Accessibility to Pick-Up Service Considering Differentiated Supply and Demand: A Case in Hangzhou, China. *Sustainability*, *11*(12), 3448.

Maere, B. D. (2017). *Ecological and Economic Impact of Automated Parcel Lockers Vs Home Delivery*. Brussels: University of Brussels.

Mansfield, E. R., & Helms, B. P. (1982). Detecting multicollinearity. *The American Statistician*, *36*(3a), 158-160.

Midi, H., Sarkar, S. K., & Rana, S. (2010). Collinearity diagnostics of binary logistic regression model. *Journal of Interdisciplinary Mathematics*, *13*(3), 253-267.

Mokhtarian, P. L. (2004). A conceptual analysis of the transportation impacts of B2C e-commerce. *Transportation*, *31*(3), 257-284.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*(1/2), 17-23.

Morganti, E., Dablanc, L., & Fortin, F. (2014a). Final deliveries for online shopping: The deployment of pickup point networks in metropolitan and suburban areas. *Research in Transportation Business & Management*, *11*, 23-31.

Morganti, E., Seidel, S., Blanquart, C., Dablanc, L., & Lenz, B. (2014b). The impact of e-commerce on final deliveries: alternative parcel delivery services in France and Germany. *Transportation Research Procedia*, *4*(0), 178-190.

Oommen, T., Baise, L. G., & Vogel, R. M. (2011). Sampling bias and class imbalance in maximum-likelihood logistic regression. *Mathematical Geosciences*, *43*(1), 99-120.

Pearce, J., & Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, *133*(3), 225-245.

Reed, P., & Brown, G. (2003). Values suitability analysis: a methodology for identifying and integrating public perceptions of ecosystem values in forest planning. *Journal of Environmental Planning and Management*, *46*(5), 643-658.

Schroeder, M. A., Lander, J., & Levine-Silverman, S. (1990). Diagnosing and dealing with multicollinearity. *Western Journal of Nursing Research*, *12*(2), 175-187.

Slabinac, M. (2015). Innovative solutions for a "Last-Mile" delivery–a European experience. *Business Logistics in Modern Management*, 111-130.

Soroush, A., Bahreininejad, A., & van den Berg, J. (2012). A hybrid customer prediction system based on multiple forward stepwise logistic regression mode. *Intelligent Data Analysis*, *16*(2), 265-278.

Squires, G. D. (2002). Urban sprawl and the uneven development of metropolitan America. In *Urban sprawl: Causes, consequences, and policy responses* (pp. 1-22). Washington, D.C: Urban Institute Press.

State Post Bureau of the People's Republic of China. Statistical Communique on the Development of Postal Industry in 2014-2019. Available online: http://www.spb.gov.cn/xw/dtxx_15079/201504/t20150429_462010.html (accessed 10.03.2021).

Taniguchi, E., & Kakimoto, T. (2003). Effects of e-commerce on metropolitan distribution and the environment. *Journal of Eastern Asia Society for Transportation Studies*, *5*, 2355-2366.

Tan, R., Xu Y., & Chen, D. (2016). Research on the spatial distribution of pickup points from the perspective of residents' behaviour: A case study of Cainiao network pickup points in Nanjing. *World Regional Studies*, *25*(5), 111-120. (in Chinese)

Weltevreden, J. W. (2008). B2c e-commerce logistics: the rise of collection-and-delivery points in The Netherlands. *International Journal of Retail & Distribution Management*, *36*(8), 638-660.

Xue, S., Li, G., Yang, L., Liu, L., Nie, Q., & Mehmood, M. S. (2019). Spatial Pattern and Influencing Factor Analysis of Attended Collection and Delivery Points in Changsha City, China. *Chinese Geographical Science*, *29*(6), 1078-1094.

Zellner, D., Keller, F., & Zellner, G. E. (2004). Variable selection in logistic regression models. *Communications in Statistics-Simulation and Computation*, *33*(3), 787-805.

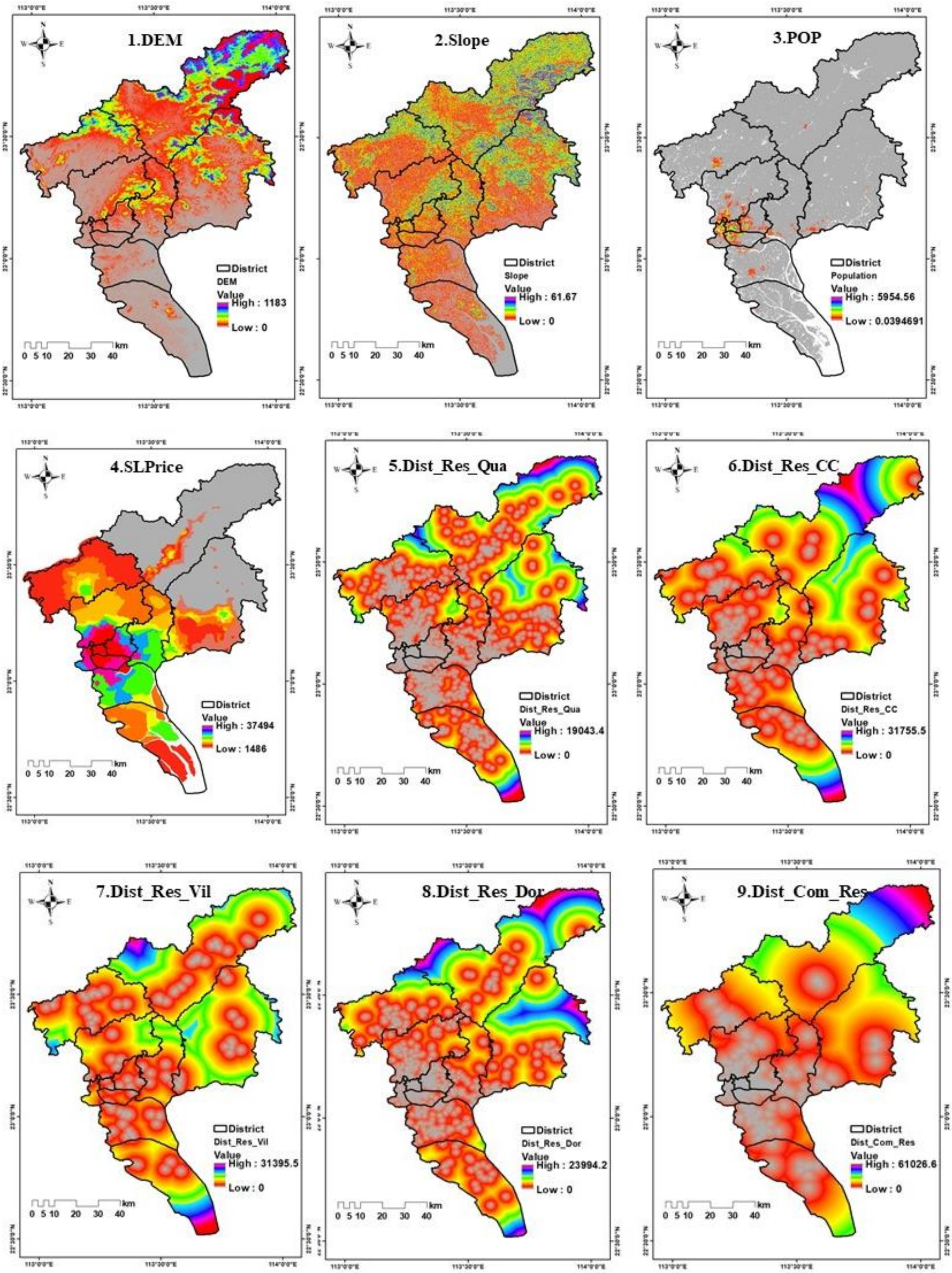
Zhang, L., Zhou, T., & Mao, C. (2019). Does the Difference in Urban Public Facility Allocation Cause Spatial Inequality in Housing Prices? Evidence from Chongqing, China. *Sustainability*, *11*(21), 6096.

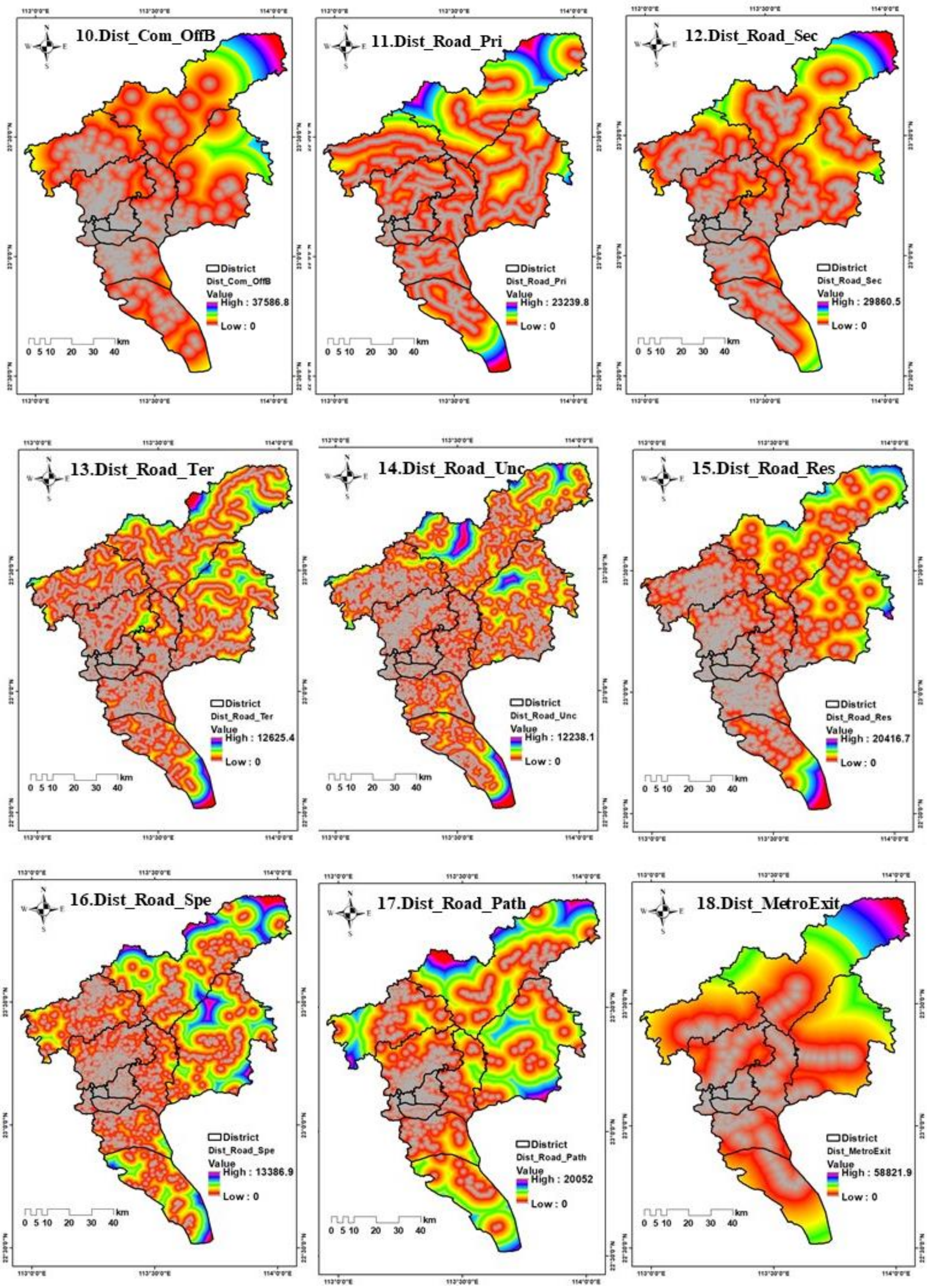
Zheng, Z., Morimoto, T., & Murayama, Y. (2020). Optimal location analysis of delivery parcel-pickup points using AHP and Network Huff Model: A case study of Shiweitang Sub-District in Guangzhou City, China. *ISPRS International Journal of Geo-Information*, *9*(4), 193.

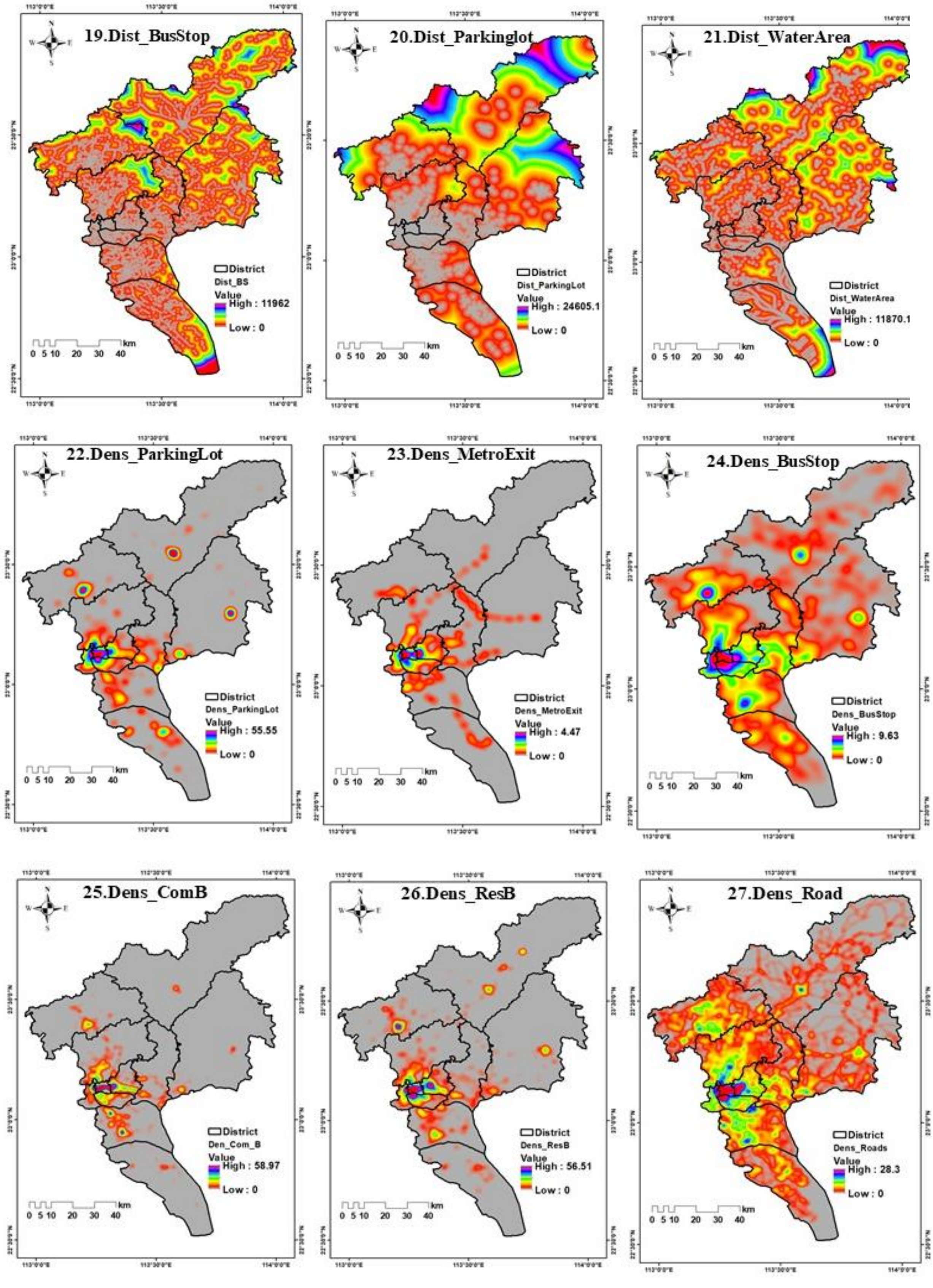
Zhou, J., Li, G., & Yang, L. (2019). Relationship Between Spatial Distribution of Cainiao Station and Metropolitan Development—A Case Study of Dongguan City. *Journal of Human Settlements in West China*, *34*(5), 83-91.

# Appendices

I. The explanatory variables for the suitability of PPPs




10.Dist_Com_OffB
District
Dist_Com_OffB
Value
High : 37586.8
Low : 0
11.Dist_Road_Pri
District
Dist_Road_Pri
Value
High : 23239.8
Low : 0
12.Dist_Road_Sec
District
Dist_Road_Sec
Value
High : 29860.5
Low : 0
13.Dist_Road_Ter
District
Dist_Road_Ter
Value
High : 12625.4
Low : 0
14.Dist_Road_Unc
District
Dist_Road_Unc
Value
High : 12238.1
Low : 0
15.Dist_Road_Res
District
Dist_Road_Res
Value
High : 20416.7
Low : 0
16.Dist_Road_Spe
District
Dist_Road_Spe
Value
High : 13386.9
Low : 0
17.Dist_Road_Path
District
Dist_Road_Path
Value
High : 20052
Low : 0
18.Dist_MetroExit
District
Dist_MetroExit
Value
High : 58821.9
Low : 0



19.Dist_BusStop
District
Dist_BS
Value
High : 11962
Low : 0
km
0 5 10 20 30 40
20.Dist_Parkinglot
District
Dist_ParkingLot
Value
High : 24605.1
Low : 0
km
0 5 10 20 30 40
21.Dist_WaterArea
District
Dist_WaterArea
Value
High : 11870.1
Low : 0
km
0 5 10 20 30 40
22.Dens_ParkingLot
District
Dens_ParkingLot
Value
High : 55.55
Low : 0
km
0 5 10 20 30 40
23.Dens_MetroExit
District
Dens_MetroExit
Value
High : 4.47
Low : 0
km
0 5 10 20 30 40
24.Dens_BusStop
District
Dens_BusStop
Value
High : 9.63
Low : 0
km
0 5 10 20 30 40
25.Dens_ComB
District
Den_Com_B
Value
High : 58.97
Low : 0
km
0 5 10 20 30 40
26.Dens_ResB
District
Dens_ResB
Value
High : 56.51
Low : 0
km
0 5 10 20 30 40
27.Dens_Road
District
Dens_Roads
Value
High : 28.3
Low : 0
km
0 5 10 20 30 40