目录

[1. Introduction 2](#_Toc476237011)

[2. Literature Review 2](#_Toc476237012)

[2.1 Catchment area 2](#_Toc476237013)

[2.2 Factors 3](#_Toc476237014)

[2.3 Model 3](#_Toc476237015)

[2.4 Summary 5](#_Toc476237016)

[3. Index Framework 5](#_Toc476237017)

[3.1 Data source 5](#_Toc476237018)

[Area of study and data 5](#_Toc476237019)

[Choice of PCA 6](#_Toc476237020)

[3.2 Variable selection and construction 6](#_Toc476237021)

[Built environment 6](#_Toc476237022)

[Transportation accessibility 7](#_Toc476237023)

[Demographic and socioeconomic environment 8](#_Toc476237024)

[4. Modeling and result 9](#_Toc476237025)

[4.1 Methodology 9](#_Toc476237026)

[4.2 Identification of Candidate Index 10](#_Toc476237027)

[4.3 Estimation of MGWR 11](#_Toc476237028)

[4.4 Residual analysis 12](#_Toc476237029)

[5. Discussion 14](#_Toc476237030)

[Built environment 14](#_Toc476237031)

[Transportation accessibility 15](#_Toc476237032)

[Demographic and social economic environment 15](#_Toc476237033)

[Limitation and implication 16](#_Toc476237034)

[Conclusion 16](#_Toc476237035)

[Reference 16](#_Toc476237036)

**Analysis of Subway Ridership at Station-Level Using Small Sample Case of Fukuoka, Japan**

# 1. Introduction

With the problem of weakness in population growth and the tendency of using private transport in the local central city of Japan, many operators of public transport in Japan, especially urban rail transit such as subway are now facing financial pressures due to the huge operating costs. How to increase the use of public transportation has become an important issue for the government of the local central city in Japan, and in order to turn around the bad financial situation [1], many efforts have begun to be taken to attract public transit user. To increase the use of public transportation, understanding how various factors affect the ridership is treated as a foundation for the policymaker. Thus, the aim of this study is proposed: exploring and explaining the factors influencing subway ridership in Fukuoka a typical local central city of Japan with a population of about 1.5 million.

From the view of methodology, exploring and estimating the factors influencing subway ridership is treated as part of traffic demand forecast[2], which is often used as part of trip generation within the four-step travel forecasting approach [3]. In the past, limited by data and analytical tools, the traditional four-step approach was usually applied to the issue of macro traffic forecast, rather than exploring and estimating the factors influencing transit ridership at station level [4]–[6]. In the four-step approach, trip generation stage is usually conducted based on an empirical model using some socioeconomic variables like population, employment, auto ownership. However, in different city cases, the impact of traffic indicators will not be exactly the same [7]. Additionally, it is also unclear whether these selected indicators are significantly related to traffic volume, or whether there is an indeed linear relationship between traffic volume and these selected indicators.

Fortunately, the development of GIS technology and the richness of digital statistics has provided a flexible framework for conducting the procedure of trip generation more accurately. Therefore, with the help of GIS technology and digital statistics, we can extract more accurate data in the catchment area of a station, thereby exploring and estimating the factors having significant relationships with ridership [8], which formed the so-called direct station-level ridership forecasting model.

This study can be viewed as an extension of existing station-level ridership model for a small sample case. Different from the case with hundreds of stations, a small sample case with dozens of stations has a higher risk of both type I and type II errors in statistic when identifying the valid variables that should be entered the regression model. To avoid such statistical errors, a procedure using exploratory regression was proposed to help to decide whether a variable should enter the regression model. The main work includes 3 aspects: 1. build the index framework based on prior study; 2. identify the valid indexes that do affect subway ridership; 3. quantitatively explain the relationship among variables in generating subway ridership using Mix Geographically Weighted Regression (MGWR) model.

# 2. Literature Review

The station-level ridership model can be considered as simplification and extension of the four-step model applied to the specific issue. The focus of this issue is almost the same with that of the four-step model, it is mainly on 3 topics summarized from the prior research: 1. How to determine the catchment area; 2. How to construct the index framework; 3. How to choose and estimate the mathematical model. *Table 1* showed the summary of literature about the issue of forecasting ridership at station level.

## 2.1 Catchment area

An important assumption for investigating factors influencing transit ridership is the definition of the catchment area of a station. The catchment area is defined as the affected area of a station, which represents the maximum distance that most people are willing to walk to the station. Because it is determined by the real walking distance of pedestrian along the street network, the catchment area is also called pedestrian catchment area (Abbreviated as PCA). Many efforts have been attempted in estimating the effective walking distance, and the thresholds were generally considered ranging from 400m to 1000m[9]–[14]. For the issue of forecasting ridership at station level, the distance threshold of 800m is the most accepted one, which is widely used as a reasonable reference value for PCA. But the 800m threshold cannot be considered as a standard, since the specific feature in urban, infrastructure and resident. Not only the physical context of urban (such as road network, bus network, walkability etc.) can have different effects in formatting PCA, but also the cultural and human environment (such as travel habit, social norm etc.) can affect the distance that people are willing to walk [15].

On this basis of the fixed distance threshold, there are also many studies finding that the greater the distance to stations within the catchment area is, the less public transit tends to be used. Therefore, some studies began to use a distance-decay function to calculate the variables within catchment area[15]. The distance-decay method now is gradually becoming accepted and it has been proved to be closer to fact. Although this distance-decay method seems to be able to provide more accuracy result, it also has some limitations, and it can’t work all the same for different study cases. One of the main reasons is, to achieve the correct distance-decay function, there must be a large-scale personal trip survey as a basis, and there are also some difficulties and limitations in implementation. In addition, because the probability of selecting transit system is hypothesized having a linear relationship with the distance between residence and station, the method of distance-decay and distance threshold is regarded as equivalent when the distribution of the population in the catchment area is not significant in spatial auto-correlation.

## 2.2 Factors

From the year of 1993, the idea of transit-oriented development (TOD) was proposed[16]. The conception of TOD has gained wide support due to the reduction of automobile dependency and improvement of urban development potential. Many research findings have shown that residents living in TOD neighborhood are more inclined to transit travel[17]–[19]. In 1997, Kockelman proposed a three dimension index system (Density, Diversity, and Design) from the perspective of TOD for this issue [20], which has been generally accepted as basic principles when analyzing transit ridership. In addition, many extensions have also been added to the 3D theory, such as accessibility to the station, connectivity of line, and capacity of station [21], [22]. In this study, all the factors expected to influence transit ridership fell into three main categories: 1. built environment; 2. transportation accessibility; 3. demographic and socioeconomic environment.

Built environment refers to the buildings or facilities that provide the setting for human activity, and it has been widely proved to have a strong relationship with ridership. Also, land use diversity has a significant effect on ridership since it reflects the balance between traffic demand and supply within the catchment area. Although the definitions of land use diversity are not the same according to different researchers, it’s widely accepted that higher diversity tends to result in less transit ridership [15], [23]–[27].

Transportation accessibility is an important factor for passengers going to take public transit. Better accessibility is thought to be attractive for passengers living further. The factors for accessibility are commonly described as the number of transfers[23], [28], network density[29], [30], number of parking facilities[28], [30] and convenience of pedestrian[4], etc. Also, the type and location of a station can affect accessibility as well. Terminal stations are more attractive for passengers because people can afford to spend more time on getting to a terminal station which is easier to transfer to other line or another mode of transportation[13].

The demographic and socioeconomic environment is also relevant to ridership. Obviously, the populations of resident and employment within the catchment area are crucial for influencing ridership of the subway station. Besides, the economic factors also played an important role in ridership. For example, in the area where the ratio of vehicle holding is higher, people will be more likely to choose private car than public transit[31], [32]; also the higher the percentage of low-income household is, the more likely people tend to take public transit[33]. Furthermore, the ratios of apartment and rental house within catchment have been verified being relevant with ridership in some degree[26].

## 2.3 Model

Ordinary least square (hereinafter abbreviated as OLS) was one of the most widely used methods for processing the issue of direct ridership model [20], [34]. However, due to the insufficient consideration of heteroscedasticity and spatial autocorrelation, the result of regression often leads to large standard errors or low level of significance. Thus, such OLS model is not available to all the stations with different characteristics. To improve the generality of the model, the model or method such as Geographically Weighted Regression (GWR), Weighted Least Squares Regression (WLS) and Poisson Regression etc. have been successively introduced into the issue of direct ridership model.

[4], [25] estimated the ridership of bus using the model of Poisson Regression. The problems occurred in ordinary linear regression such as the contravention between fact and estimated coefficients, and the low level of significance was well addressed. Therefore, both generality and explanatory ability in the regression model were enhanced.

The study of [28] has achieved a high coefficient of determination at the first stage with OLS, however, the result showed a significant heteroscedasticity and a non-random distribution of estimated residual. To deal with this deviation at the second stage, WLS was brought in to eliminate heteroscedasticity, in which the data points were weighted using the standard error. The result showed that WLS was effective in eliminating heteroscedasticity and improving the explanatory ability of the model.

Table 1

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | | 2004 | 2004 | 2009 | 2010 | 2011 | 2012 | 2013 | 2013 | 2015 |
| Author | | Chu | Kuby et al. | Taylor et al. | Sohn and Shim | Gutiérrez et al. | Cardozo et al. | Chakraborty et al. | Zhao et al. | Jun et al. |
| Catchment | | ¼ mile (400m) walking distance | Half mile (800m) walking distance | N/A | N/A | Distance-decay 800m buffer | 800m walking distance | N/A | 800m radius | 300m,600m,900m radius |
| Method | | Poisson Regression | WLS | 2SLS | OLS/SEM | OLS | OLS/GWR | OLS/SEM | OLS | OLS/MGWR |
| Sample Size | | 2568 | 268 | 265 | 251 | 158 | 190 | 900 | 55 | 442 |
| Number of Valid Indicator | | 15 | 11 | 8 | 7 | 9 | 4 | 9 | 11 | 11 |
| Coefficient of determination (Adjusted R2) | | 0.54 | 0.71 | 0.91 | 0.60 | 0.73 | 0.56 | 0.69 | 0.95 | 0.77 |
| Building environment | Built environment |  |  |  | ● | ● |  | ● | ● | ● |
| Hospital |  |  |  |  |  |  |  | ● |  |
| School/University |  |  |  | ● |  |  |  | ● |  |
| CBD |  | ● |  |  |  |  |  | ● |  |
| Land use mix |  |  |  | ● | ● | ● |  |  | ● |
| Other infrastructures |  | ● |  |  |  |  |  | ● | ● |
| Transportation Accessibility | Accessibility of pedestrian | ● |  |  |  |  |  | ● |  |  |
| Accessibility of transfer |  | ● |  | ● | ● | ● | ● | ● | ● |
| Road coverage |  |  | ● | ● |  |  |  | ● |  |
| Parking |  | ● |  |  |  |  |  | ● |  |
| Service level of public transit | ● | ● | ● |  | ● | ● |  | ● | ● |
| Locational factor |  | ● |  | ● |  |  |  | ● |  |
| Demographic and Socioeconomic Environment | Population |  | ● | ● | ● |  | ● | ● | ● | ● |
| Employment | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| Age | ● |  | ● |  |  |  |  |  | ● |
| Tenant proportion |  | ● |  |  |  |  |  |  | ● |
| Race | ● | ● | ● |  | ● |  |  |  |  |
| Income | ● |  |  |  |  |  | ● |  |  |
| Vehicles holdings | ● |  |  |  |  | ● | ● |  |  |
| Fare |  |  | ● |  |  |  |  |  |  |

The data points in OLS are regarded to be independent of each other, however, each data point has different geographical location in the issue of direct ridership model, the observed values are not considered to be independent of each other in terms of the fact that they are distributed continuously in space. For one data point in regression, the observed value is related to the data point nearby in geographical location, and the regression parameters in different geographical locations usually have different performances in their characteristics[35]. For the problems of spatial autocorrelation and spatial heterogeneity, [24] made a comparison of OLS and GWR with the same regression parameters, the result showed that the coefficient of determination had a significant improvement and the standard errors turned to be less in GWR. On the basis of common GWR, [26], [31] introduced Mixed Geographically Weighted Regression (MGWR) to this issue in the consideration of that some regression parameters didn’t have special autocorrelation. They set part of the parameters as global independent variables, and the others as spatially autocorrelated variables, to make model closer to fact.

## 2.4 Summary

Although existing studies have done a lot on the issue of direct ridership model, there is still some insufficiency in each study due to the limitation of study case and data source. For the selection and construction of factors, the simple and direct indicators such as population, employment etc. are roughly the same with existing studies. But the definition of indicators obtained by secondary calculating such as land use mix, bus service etc. are not the same, and the effects of such factors have not been well verified and widely accepted. For the model, most of the coefficient of determination in OLS were not very ideal (less than 0.7), there was still more than 30% of the change in ridership not being explained by the model [15], [26]. Additionally, even though some of the studies have obtained the high coefficient of determination exactly, there was still another problem that one factor had too strong effect while the rest of the factors had very little influence on the ridership.

This study attempts to address some of the shortcomings existed in the previous studies. First, optimize the index framework, and proposes some new indicators to help describe the variety of ridership. Second, optimize the procedure of identifying a valid explanatory variable, especially for a small sample case. Third, propose an approach for identifying if an indicator a global or local in MGWR model, to prevent a small sample case from becoming into data-driving.

# 3. Index Framework

## 3.1 Data source

### Area of study and data

This study focuses on 35 subway stations in Fukuoka City (the sixth largest city in Japan) which has the largest population in Kyushu Island of Japan (more than 1.5 million). Figure 1 is the research area and the distribution of subway stations. Until now Fukuoka has three operating subway lines (The first line was operated since 1981, the second line was operated in 1993 and the latest third line began in 2005), a total of 29.8 kilometers operating mileage. The transport system carries a daily average of more than 0.4 million passengers by 2015 that accounting for more than 20% in total motorized travel[36]. Although the subway system of Fukuoka is not a large-scale one, it plays a crucial role in public transportation in terms of the city scale and population.

Most of the data used in this study can be freely downloaded or bought from the government official website, except the data of Basic Survey of City Planning which is obtained from the Department of Transportation in Fukuoka. The data of H22 is used as a reference. All the resource of data used in the study is listed below.

- subway ridership

- Basic Survey of City Planning

- Census

- Digital Map (Basic Geospatial Information)

- National Land Numerical Information

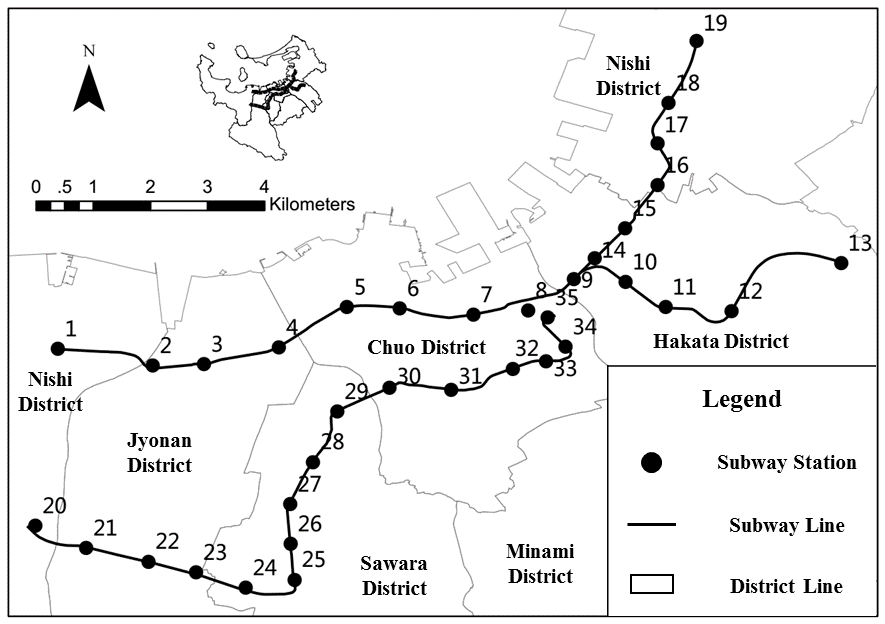


Figure 1

### Choice of PCA

At present, in the United States, a half-mile-radius (800 m) circle has become the practical standard for rail-transit catchment areas based on TODs [37]. The distance of 800m corresponds to the distance people can walk in 10 minutes at the speed of 4.8 km/h. A Japanese case study [38] also supported this 800m catchment area for TOD by using the survey data of <2010 big city traffic census metropolitan area report>.

Based on the report of basic survey along the subway third line in Fukuoka, more than 70% of the passengers choose to walk to the station, about 16% choose bicycles, the total percentage of non-motorized trip accessing to station accounted for about 90%. The time for walking is 7.7 minutes on average by the speed of 4.8km/h, which means the average walking distance is about 600m for walking to the station. Considered that the average walking distance should be a little less than the main walking distance, it can be inferred that the PCA in Fukuoka is consistent with the most widely accepted distance of 800m.

Based on the summary of existing literature and the condition of Fukuoka, the distance threshold of 800m is adopted in this study. All the data based on geographical information will be covered by the 800m PCA using the areal interpolation method.

## 3.2 Variable selection and construction

In this study, the average daily ridership of each subway station is used as dependent variable. The candidate independent variables are shown in *Table 3*. Based on the previous studies, the influence on ridership can be summarized into 3 aspects in this study: 1. built environment; 2. transportation accessibility; 3. demographic and socioeconomic environment.

There are 16 variables in total adopted as candidate factors influencing ridership. As is known that some of these variables have already been estimated many times in previous studies, however, there are still some indicators, which may help explain the variation of ridership, ignored in the prior studies. Besides, there is another problem to deal with that the more comprehensive the index framework is, the problem of multicollinearity will occur more easily. To deal with this problem, not all the candidate variables will enter the regression model, they should be identified and selected by interpretability in the next stage.

### Built environment

The building provides people space for living, working and recreating, and people tend to have different travel preferences with different trip purposes. Therefore, the floor area is commonly thought to be the primary factor that can affect ridership. Another compound indicator, Land use mix, is also thought to be an important factor that can influence transit ridership because people living in the catchment area with higher land use mix can do most of their daily activities at different types of building without taking public transport. Although the index (Eq. 1) of land use mix introduced by [39] was commonly used in other studies, it cannot reflect the mixture accurately because of the equal treatment of all types of land use.

Eq. 1

Where is the indicator of land use mix, is the total floor area within catchment area, is the number of all land use types being considered, is the floor area of land use type .

The same with previous studies, this study chose several types of land use with high proportion to assess the indicator of land use mix, including residential, office, commercial, education. The four main types of land use account for about 90% of all the floor area in Fukuoka City, especially in subway catchment area, reaching more than 95%. However, different from the general definition of land use mix, the proportion of land use is not evenly distributed in the case of Fukuoka City. As shown in *Table 2*, obviously, it is significantly different from average distribution, and the proportion varies with the variation of the catchment area. To describe the mixture of land use closer to the facts, the referenced balance proportion of land use types is decided by the average proportion of all subway station catchment area (800m pedestrian distance) in Fukuoka City. In addition, to make indicator more intuitive and simpler, the index of land use mix is redefined as the aggregation of land use. The Euclidean Metric is used for evaluating the deviation of land use aggregation in each subway station with respect to a reference value. The value of this indicator is arranged from 0 to 0.5, in which the lower value represents a higher level of mixing in land use function, while the higher value means the land use function is more single-minded. This indicator of land use aggregation is defined as Eq. 2, it is speculated to have a negative impact on ridership.

Eq. 2

Where is the indicator for aggregation of land use functions, is the average proportion of type land use, represents the type of land use (respectively government, commercial, residence and education), is the floor area of land use type within catchment area of subway station.

Table 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Business | Commerce | Residence | Education |
| 600 m Pedestrian distance | 20.8% | 14.9% | 55.7% | 6.0% |
| 800 m Pedestrian distance | 18.8% | 12.8% | 59.4% | 6.7% |
| 1000 m Pedestrian distance | 17.7% | 11.6% | 62.0% | 6.5% |
| Fukuoka City | 10.5% | 7.7% | 65.2% | 5.5% |

### Transportation accessibility

Transportation accessibility represents the convenience and cost to arrive subway station, it can affect ridership in two opposite ways. Higher accessibility of catchment area means people will have more transportation options to get into the area instead of the subway, such as private cars, bicycle, bus or walking etc. Also, higher accessibility allows people to arrive station more easily, and then lead to an increase in ridership. Considering different travel modes, accessibility is separated into 3 parts to be interpreted in this study. First is for users of the road network; second is for passengers transferring from the bus; third is for rail transit interchange.

Road network is commonly used in previous studies. Users of the road network are not only the vehicle but also non-motorized traveler including bicycle rider and pedestrian. As is known, the higher coverage of road network is thought to have better accessibility for both motor travel and non-motorized travel. Additional, bicycle parking may also be an important factor influencing the accessibility of catchment area for non-motorized travelers. Therefore, the accessibility index for road network is set into the number of bicycle parking and road density in this study.

Another accessibility for passenger transferring from the bus is also considered from both positive and negative effects. Bus service in catchment area can reflect the connectivity with other regions, part of the potential subway ridership can be shared by bus. This sharing ridership can be represented by the service capacity of all the bus operating in the same catchment area of the subway station, and the service capacity () is expressed by Eq. 3 as follow. In contrast, the bus station close to subway can be used as transfer station which can increase the accessibility of subway for people who want to take the subway but living far away from subway station, thus having a positive effect on ridership of subway. The indicator should represent the accessibility between subway and bus, thus it is defined as the number of bus lines within a smaller area around the center of the subway station. The indicator representing accessibility of bus is defined as Eq. 4 in the below.

Eq. 3

Where is the number of bus stations within catchment area of subway station, refers to the number of lines in one bus station and is the frequency of NO. line at NO. station.

Eq. 4

Where the is the number of bus stations within catchment area of subway station, is the number of bus lines passing through the th bus station.

Because the existing rail transit network is relatively stable and unchanging, the convenience of railway interchange is roughly determined by the characteristic of the subway station. The accessibilities are different in diverse types of station: intermediate, terminal, interchange, intermodal.

Terminal station is generally thought to have a larger catchment area since the terminal station is probably the only choice for the people living far away from the station when they want to take the subway. And people can accept spending more time on the way to the terminal station [13]. Interchange station and intermodal station are attractive for passengers since it can connect to other lines of railway transit or other modes of transportation [28]. To distinguish the difference between different types of stations, the dummy variable for describing the number of railway lines passing through each station is introduced into this study.

### Demographic and socioeconomic environment

The same with previous studies, variables on population and employment should be the first to consider. Age structure and household member are also considered to affect travel habit. Besides, the index of job-resident balance may be one of the determinants in estimating internal travel within the catchment area. Generally, it is expected that the family with more people at work tends to generate more travel. The tenant proportion was hypothesized to be negatively connected to station boarding because renters are usually commute-oriented, they prefer to live close to where they are working and usually commute by walking.

Table 3

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Category | Variable | Expected sign | Min Value | Max Value | Average | Unit |
| Built environment | Commerce | + | 2,921 | 811,281 | 114,353 | m2 |
| Office | + | 2,614 | 839,956 | 167,088 | m2 |
| Residence | + | 110,748 | 1,067,523 | 528,533 | m2 |
| Education | + | 294 | 305,559 | 59,691 | m2 |
| Government | + | - | 128,471 | 20,878 | m2 |
| Transportation Facility Area | + | 197 | 132,777 | 21,204 | m2 |
| Land use Aggregation | + | 0.09 | 0.75 | 0.31 | - |
| Transportation Accessibility | Transfer Dummy | + | 1 | 4 | 1.34 | - |
| Bicycle Parking | Unknown | 64 | 4,375 | 778 | - |
| Bus Capacity | Unknown | 3 | 260 | 58.48 | - |
| Bus Accessibility | Unknown | 4 | 455 | 89.71 | - |
| Road Density | - | 191 | 479 | 299 | m/ha2 |
| Demographic and Socioeconomic Environment | Population | + | 1,908 | 19,393 | 9,813 | - |
| House Member | - | 1.86 | 2.79 | 2.18 | - |
| Job-Resident Balance | Unknown | 1.27 | 2.61 | 1.80 | - |
| Tenant Proportion | - | 0.15 | 0.65 | 0.43 | Percentage |

The statistical description of all 16 candidate indicators are shown in *Table 3*, most of the indicators have been proved in previous studies, based on which an expected influence is shown in the column of expected sign. However, there are also some indicators not proved in previous studies such as bicycle parking, bus capacity, bus accessibility, and job-resident balance. Whether they have a significant influence on subway ridership and how they affect the subway ridership in Fukuoka City are verified below.

# 4. Modeling and result

## 4.1 Methodology

Based on the index framework summarized above, how these indicators can influence the subway ridership is estimated as follow. With the consideration of small sample case in this study, the method is divided into two phases, identifying valid variables and estimating coefficient.

The exploratory regression tool in ArcGIS is introduced into this study to help to conduct the process of identifying valid variables. As is known, finding a proper OLS model is the main problem in this kind of study, especially when there are lots of candidate explanatory variables planned to be estimated in the regression model. The Exploratory Regression tool provides a reference for choosing a valid combination of explanatory variables. It is a data mining tool that will try all possible combinations of explanatory variables to see which model can pass the necessary OLS diagnostics. This study proposes a two-step procedure to explores the final model with an optimal combination of explanatory indicators, rather than selecting the best one from all possible combinations. The first step is selecting the variables having effectiveness in explaining the subway ridership, and the second step is choosing the best combination of the valid variables using as final indicators.

GWR is a spatial regression technique, which is used for dealing with the explanatory variables with spatial dependence. The coefficients of explanatory variables are varied with the spatial location of the data point in GWR, and the closer the distance between a data point and observation point is, the greater weight the data point is. Different from general GWR, the explanatory variables in MGWR can be either spatial dependent or spatial independent. The variables with spatial dependence (called local variable) are the same with that in GWR, varied with spatial location of data points; while the variables without spatial dependence (called global variable) are the same with that in OLS, constant in all data points. Before estimating the MGWR model, it’s necessary to determine whether the variable is spatial dependent or not. To prevent this small sample case from becoming data-driven, repeating test is conducted to reduce the probability of occasional mistake. The local/global variables were determined by the spatial dependency of each exploratory variable, in which the variable with spatial dependency is treated as a local term, otherwise, is treated as a global term.

## 4.2 Identification of Candidate Index

The first stage of selecting effective variables is conducted based on three judgment factors: 1. Multicollinearity, which is expressed by the factor of VIF; 2. Validity, which is expressed by the number of times that shows statistical importance; 3. Stability, which is shown by the percentage of negative and positive effect to the dependent variable (*Table 4*). Generally, the variable is thought to be multicollinearity if the value of VIF factor more than 7.5, as shown in *Table 4* there are four variables with higher VIF which are marked by dark color. And Figure 2 shows the probability of simultaneous appearance on the multicollinearity. For the factor of validity, three variables are filtered since they rarely show their statistical significance (less than 10%). Even though some variables have statistical significances in the regression model, they are still not credible since their performance are not stable in different models (sometimes they are positive for the independent variable but sometimes are not). As shown in *Table 4*, the three variables marked with dark color are shown not stable in explaining dependent variable. Therefore, 10 valid variables of 16 candidate variables are kept in the first stage.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 4 | | | | | | | | | | |
| Category | Variable | All Entry by OLS | |  | Statistical Information of Exploratory Regression | | | | | |
| B | Sig. |  | VIF | Validity | stability | times | + | - |
| Total |  |  |  |  |  |  | 367 | 253 | 114 |
| Built environment | Commerce Area | .003 | .801 |  | 8.8 | 32.1% | 100.0% | 25 | 25 | 0 |
| Office Area | -.005 | .584 |  | 10.4 | 9.0% | 71.4% | 7 | 5 | 2 |
| Residence Area | .017 | .168 |  | 27.5 | 12.8% | 70.0% | 10 | 7 | 3 |
| Education Area | -.002 | .897 |  | 1.8 | 7.7% | 100.0% | 6 | 6 | 0 |
| Government Area | .042 | .079 |  | 1.8 | 42.3% | 100.0% | 33 | 33 | 0 |
| Transportation Facility Area | .108 | .015 |  | 3.0 | 67.9% | 100.0% | 53 | 53 | 0 |
| Land use Aggregation | 17913.483 | .022 |  | 2.5 | 50.0% | 100.0% | 39 | 39 | 0 |
| Transportation Accessibility | Transfer Dummy | 7460.118 | .000 |  | 3.0 | 34.6% | 100.0% | 27 | 27 | 0 |
| Bicycle Parking | 6.580 | .000 |  | 4.5 | 44.9% | 100.0% | 35 | 35 | 0 |
| Bus Capacity | -75.359 | .029 |  | 6.6 | 37.2% | 96.6% | 29 | 1 | 28 |
| Bus Accessibility | 63.026 | .008 |  | 6.8 | 19.2% | 100.0% | 15 | 15 | 0 |
| Road Density | .202 | .168 |  | 2.3 | 1.3% | 100.0% | 1 | 1 | 0 |
| Demographic and Socioeconomic Environment | Population | -.762 | .297 |  | 23.9 | 12.8% | 60.0% | 10 | 6 | 4 |
| Household Members | -2783.848 | .415 |  | 3.0 | 39.7% | 100.0% | 31 | 0 | 31 |
| Job-Resident Balance | -5396.131 | .083 |  | 3.0 | 37.2% | 100.0% | 29 | 0 | 29 |
| Tenant Proportion | -5078.346 | .486 |  | 2.2 | 21.8% | 100.0% | 17 | 0 | 17 |

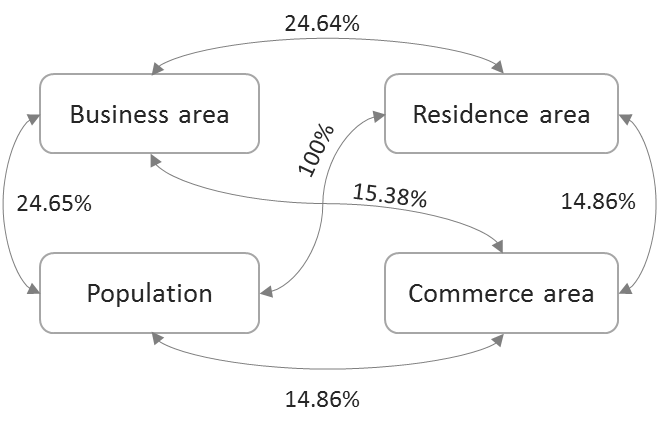


Figure 2

At the second stage, the exploratory regression is conducted again to select an optimal combination of explanatory variables based on a statistical test for regression. As shown in *Table 5*, there are 9 valid variables entering the model at last (at 95% confidence level). The JB test is not significant in the final model; it indicates that there are no biased standard errors due to heteroscedasticity. The K (BP) test is not showing statistical significance, it represents that the residuals are not deviating from a normal theoretical distribution. The test of SA is not significant; it means the residuals are not spatial autocorrelated. The optimal combination with 9 explanatory variables will be evaluated by using MGWR in the next part to obtain a better result with fewer residuals.

Table 5

|  |  |  |  |
| --- | --- | --- | --- |
| Independent variables | Test model | | |
| Beta | Sig | VIF |
| Government Area | 0.10 | 0.02 | 1.36 |
| Transportation Facility Area | 0.20 | 0.00 | 2.31 |
| Land use Aggregation | 0.10 | 0.03 | 1.42 |
| Bicycle Parking | 0.46 | 0.00 | 2.60 |
| Bus Capacity | -0.20 | 0.01 | 3.56 |
| Bus Accessibility | 0.26 | 0.00 | 4.71 |
| Transfer Dummy | 0.26 | 0.00 | 2.79 |
| Job-Resident Balance | -0.10 | 0.05 | 1.94 |
| Tenant Proportion | -0.09 | 0.03 | 1.32 |
| Household Members | - | - | - |
| Residual sum of squares | 337744990 | | |
| Adjusted *R2* | 0.96 | | |
| AICc | 694.39 | | |
| Jarque-Bera test (Sig) | 0.61 | | |
| Koenker (BP) test (Sig) | 0.85 | | |
| Spatial Autocorrelation test (Sig) | 0.27 | | |

## 4.3 Estimation of MGWR

The MGWR is estimated using the GWR4 software. The determination of local or global is processed in two steps: firstly, using Moran’s index to examine if the variable is spatial autocorrelation or not; secondly, re-examining the variable with spatial dependency in terms of the indicator of “DIFF od Criterion” provided by GWR4. The test value which is describing the spatial relationship is summarized in Table 6. Five variables are found to be spatial autocorrelation; they are transportation warehousing, bus capacity, bus accessibility and the tenant proportion respectively. Thereby in the first step, the other 4 variables are considered as global variables while the 5 variables with spatial dependency are considered as local variables. Nakaya [40], the author of the GWR4 software suggested in the GWR4 user manual that the assessment of the spatial variability of the kth varying coefficient is conducted by comparing the fitted GWR with a model in which only the kth coefficient is fixed while the other coefficients vary spatially. If the original model is better than the model with the kth coefficient fixed, that coefficient can be considered as spatial autocorrelation. GWR4 also provides the indicator of model comparison which is “DIFF of Criterion”. The user manual suggested that a positive “DIFF of Criterion”, especially greater than or equal to two, means the local term is better to be assumed as global. As shown in Table 6, all the indicator “DIFF of Criterion” are no more than two, which means that all the local terms are adapt to the model. Therefore, all the local variables have passed the test in the second step.

Table 6

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Moran's Index | z-score | p-value | Pattern | DIFF of Criterion | Type |
| Government Area | 0.04 | 0.66 | 0.51 | Random | - | Global |
| Transportation Facility Area | 0.29 | 3.33 | 0.00 | Clustered | -1.95 | Local |
| Land use Aggregation | -0.01 | 0.20 | 0.84 | Random | - | Global |
| Transfer Dummy | 0.13 | 1.58 | 0.12 | Random | - | Global |
| Bicycle Parking | -0.12 | -0.92 | 0.36 | Random | - | Global |
| Bus Capacity | 0.70 | 7.57 | 0.00 | Clustered | 0.18 | Local |
| Bus Accessibility | 0.45 | 5.04 | 0.00 | Clustered | 0.04 | Local |
| Job-Resident Balance | 0.77 | 7.61 | 0.00 | Clustered | -0.17 | Local |
| Tenant Proportion | 0.24 | 2.54 | 0.01 | Clustered | 1.02 | Local |

The MGWR model is estimated by using a fixed Gaussian kernel function and using a “golden-section search” method to find the optimal bandwidth size. The optimal bandwidth is obtained when its corresponding AICc value gets to the minimum. The variation of AICc at different bandwidths are shown in Figure 3, and the bandwidth for the lowest AICc is 5.7km.

Figure 3

Table 7 reports the result of MGWR model. First, the global variables in MGWR model have statistical significance at 95% confidence level, it can be considered reliable in statistics. The signs of all terms for local variables are consistent with that of OLS model in Table 4, and the values of coefficients in MGWR maintain a high consistency with OLS model. Moreover, comparing with the result of OLS regression, the result of MGWR has an improvement in both adjusted *R2* and AICc value, and there is a 12% decrease in residual.

Table 7

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | MGWR model | | |
| Variable | Type | Coefficient | SE | t |
| Government Area | Global | 0.05 | 0.02 | 2.59 |
| Transportation Facility Area | Local | 0.1 | 0.02 | - |
| Land use Aggregation | Global | 13,384.34 | 5,408.81 | 2.48 |
| Transfer Dummy | Global | 5,968.65 | 1,198.72 | 4.98 |
| Bicycle Parking | Global | 7.72 | 0.9 | 8.59 |
| Bus Capacity | Local | -55.14 | 5.94 | - |
| Bus Accessibility | Local | 48.61 | 2.43 | - |
| Job-Resident Balance | Local | -2,411.17 | 364.44 | - |
| Tenant Proportion | Local | -10,304.66 | 756.09 | - |
|  |  |  | | |
| Best bandwidth |  | 5.7km | | |
| AICc |  | 690.60 | | |
| Residual sum of squares |  | 296311499 | | |

## 4.4 Residual analysis

As the result from both models, MGWR has a better performance than OLS, in which the residual of MGWR is 12.27% less than that of OLS. Also, the homogeneity of residual distribution is evaluated by using Moran’s index shown in *Table 8*. As can be seen, the Moran’s index in MGWR is closer to expected value than that in OLS. The MGWR model also shows less variance and greater likelihood of random distribution (with lower z-score and higher p-value) than OLS model. The spatial autocorrelation analysis presents that the residual of OLS model is more likely to be aggregative in space.

Table 8

|  |  |  |
| --- | --- | --- |
|  | OLS | MGWR |
| Moran’s index | 0.08 | 0.03 |
| Expected index | -0.03 | -0.03 |
| Variance | 0.01 | 0.01 |
| z-score | 1.09 | 0.61 |
| p-value | 0.27 | 0.54 |

Since the MGWR is a kind of variable coefficient regression model, the coefficients and adjusted *R2* of each data point are different depending on the location of the data point. For the 5 local variables in all 35 data points, the spatial distribution of significance and local *R2*is mapped in Figure 4. For the variable of building area of transportation, bus capacity, bus accessibility, job-resident balance and tenant proportion, there are 33, 34, 35, 24 and 32 data points having statistical significance at confidence level 95% respectively. As can be seen, the two indicators for bus proposed in this study have good performance in terms of significance and stability, while the variable of ‘tenant proportion’ doesn’t have a strong stability in explanatory ability. Relatively, the other 4 variables have higher reliability in explaining the variety of ridership.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  | |
|  | |

Figure 4

# 5. Discussion

This study focuses on small sample size, using 9 explanatory variables to describe the variety of subway ridership in a local central city with 35 subway stations, while regarding the outcomes of some existed studies in Table 1, most of these studies have a middle sample size (ranged from 150-450), and the number of explanatory variables is general about 10. The adjusted *R2* (0.96) in this study is a relatively higher one comparing with other studies. Most of the variation in ridership can be explained by the 9 variables in this present model. But the adjusted *R2* is not the only judgment criterion in the regression model, especially in the small sample case. The result of this study shows that there are 3 variables, which are bus capacity, job-resident balance, and tenant proportion, impacting negative effect on ridership, while the others 6 variables plays positive effect. It is interesting to note from the coefficient that although the variables of bus capacity and bus accessibility have a strong positive correlation with statistical significance, they perform fully opposite effect on the independent variable. Regarding the coefficient showed in Table 7 obtained from the two models and 9 variables, the coefficients in both OLS model and MGWR model have consistent signs and similar values. The rationale of estimated results will be discussed from the three categories of indicators.

### Built environment

Built environment has been considered as a critical driver for influencing transit ridership[23], and some results of empirical studies have supported it [34], [41]. Built environment can reflect land use density around the station, different kind of building function has different demand in using subway [34]. In this study, two variables about building floor area (government area and transportation facility area) are found to be valid in explaining subway ridership (with statistical significance at 0.05 level). The coefficients showed in Table 7 implies that there is an increase of 5/10 passengers per additional 100 m2 building floor area for government/transportation facility respectively.

Refer to the results of previous studies, the building floor area of office and commerce was normally considered as the crucial driver for generating transit ridership. However, travel habit and culture are not the same in all cities. The result of this study stated that in Fukuoka people working at or visiting government office are likely to use the subway. Another valid indicator is building floor area of transportation facilities, which mainly represents the scale of public transit in the urban area. Obviously, the larger station usually has a larger scale of passengers, but the causal relationship is that forecasted ridership determined the scale of the station, rather than the opposite direction. Thus, this indicator of transportation facility can be viewed as an index for posterior evaluation, to judge if the planning is consistent with the fact, but not a predictable index. In fact, the variables of office area, commerce area, and residence area are also placed on the candidate list, but they are not showing statistical significance. One possible conjecture is given here: these three indicators represent the major category in building type, but they also contain several subcategories which cannot be expressed in the indicators. It means that each of these indicators is interpreting multiple issues, for which they cannot be consistent with statistics.

Moreover, the land use mix is also widely thought to be a crucial factor that can influence subway ridership. [15], [26] argued that the diversity of land use has a positive relationship with ridership, which means the higher diversity of land use can attract more passengers. However, the finding of this study shows an opposite result. The index of land use mix is redefined into land use aggregation in this study, and the result shows that the more aggregated the land use is the more ridership will be generated. In another word, the high diversity of land use will lead to a decrease in subway ridership. The relationship can be interpreted as following from the perspective of principles and features of TOD. A complete TOD area should have various kinds of urban function, and the main aim of TOD is to allow people to do their daily activities by walking in the TOD area, thus reducing inefficient vehicle trip. That is, if people can do most of their daily activities around the station, they will tend to reduce the use of subway. It is still not clear why this difference occurs. And unfortunately, the indicator of land use mix was not discussed in previous studies, there is little reference to explain it. A hypothesis is proposed to interpret this difference, even though there is no way to verify it. Regarding the generally used definition of this indicator in previous studies (*Eq. 1*), which was proposed by [39] in 2007, this equation treated each kind of land use equally. But in fact, the proportion of each kind of land use should not be equal, the proportion in Fukuoka is as shown in Table 2. Therefore, maybe this indicator with statistical significance in the prior studies was interpreting something else related to ridership but not describing land use mix. The results may be just statistically relevant to the ridership coincidentally. And that is the reason why this study proposes the method of identifying valid explanatory variables. Especially for small sample case, repeating test can reduce the probability of contingency.

### Transportation accessibility

Accessibility is also thought to be a key factor influencing ridership. [23] further divided this factor into internal and external accessibility, which was also cited and used by [41], the former represented the accessibility to station within the catchment area, and the latter reflected the connectivity to the place outside the catchment area. This study identified four valid variables in transportation accessibility: transfer dummy, bicycle parking, bus capacity and bus accessibility. The transfer dummy and bicycle parking can be easily classified into external accessibility and internal accessibility respectively, and both have a positive effect on ridership. It means the more easily people reach subway station, the more likely they will use the subway. This result is also consistent with prior studies [15], [24], [28].

However, the effect of bus service on subway ridership is not so intuitive and clear as other factors. It can be speculated that bus service may have both positive and negative effects on ridership, for there are both competitive and transferring relationships between bus and subway simultaneously. Therefore, a greater transport capacity of bus service can share part of the passengers of the subway while a more accessible route network, bus service can transfer more passengers to subway from other places. And the result from this study has verified this hypothesis that the indicator of bus accessibility and bus capacity showed the totally opposite effect on subway ridership, the former have a positive effect on subway ridership while the latter is in contrary. The indicator of bus service was intuitively considered to be related with subway ridership, and it often appears in the candidate variables list of prior studies, but is rarely estimated successfully in final model (some studies used the indicator of feeder bus [23], [24], [41] rather than normal urban bus. Besides, [23] also considered the factor of trunk bus which is thought to be competitive with subway, but it did not show statistical significance). It can be guessed that the influence of bus service cannot be interpreted by only one indicator since the factor of bus service contains more than one kind of information. This result provided some inspiration that every transportation mode having both competitive and transferring relationships with another transportation mode may have both positive and negative effect on the others simultaneously.

### Demographic and social economic environment

Regarding the factor of the demographic and socioeconomic environment, the job-resident balance and tenant proportion are found to be effective in explaining the variation of subway ridership. As shown in Table 7, both job-resident balance and tenant proportion are showed to have a negative impact on subway ridership. The result indicates that working people tend to use subway more than unemployed people (like children, old people, and housewife), while tenant takes subway less.

Because the travel habit is not the same between working people and unemployed people, the indicator of job-resident balance is developed to help to interpret the difference between jobs and population. [42] suggests that job-resident balance is a crucial factor that can influence house price, this indicator is thought to be related to income level, family structure, and social class etc. But whether and how it can influence subway ridership has not been verified yet in prior studies. The result from this study shows that job-resident balance can affect subway ridership, and it also can be inferred that different group of people has different travel habit. Moreover, the generally considered crucial indicator of population and employment is not showing validity in this study. One possible explanation is that these two variables have multicollinearity with other variables and they have been expressed in the combination of other variables.

In this study, tenant proportion is also verified that having a positive influence on subway ridership. But for different cities and regions, travel preferences are different. The result from the empirical case in nine US cities indicates that tenant is more likely to use subway [28], while the case in Seoul shows an opposite result that tenants living around station use subway less[26]. It seems that renters are likely to use public transport, since they are thought to be poor, young, located in denser multifamily housing [28]. However, the discussion also suggests that the indicator of tenant percentage should be treated separated: it may have a high tenant percentage in both CBD areas and suburban apartment, but of which the travel habits may be totally different. That means even though travel preference is different in CBD area and suburban area, the indicator of tenant proportion may be almost the same.

### Limitation and implication

This study focuses on subway ridership at station level. However, the ridership of one station is not only influenced by the circumstance of itself but also the stations connected to it. Once the circumstance within catchment area changes, obviously, the ridership of this station will change as well. But the increased part of passengers will be transported to other stations, thus every station connected to that station will also have an increasing on ridership. Until now, most of the studies (including this study) on this issue are focusing on station level. Therefore, identifying and explaining the factors influencing subway ridership from both origin and destination will be the next step of research. Meanwhile, the analysis of subway ridership at station-to-station level based on OD data will also provide a way to research the spatial connectivity among the catchment area of subway stations.

### Conclusion

This study examined the factors may be associated with transit ridership using the case of subway stations in Fukuoka of Japan. 9 effective factors were selected from candidate indicators to describe the variation of subway ridership in the final models. As discussed above, the result can be considered reasonable based on logical common sense and previous studies. The major contribution of this study can be summarized as follows.

First, on the base of previous studies, this study reclassified and reorganized the indicator framework. Besides the indicator appeared in the previous studies, the indicator of land-use mix was redefined as an aggregation of land use to make the expression of this index more intuitive and closer to reality. Additionally, two variables representing bus accessibility and bus capacity were proposed in this study to explore how can bus service influence the subway ridership.

Second, the approach of selecting variables from candidates, which was conducted using the tool of exploratory regression in ArcGIS, was proposed and proved to have a good performance in dealing with a small sample case. Comparing with the approach that selecting valid variables by choosing a combination with the highest Adjusted R2 among all the test models, the approach used in this study could avoid the occurrence of accidental errors to some extent since it chose the valid variables based on the statistics obtained from repeating the trial regression.

Third, how to determine whether a variable is global or local has always been a problem in MGWR, although the tool of GWR4 has provided an approach to address this issue, it is still a data-driven approach method which may not be applicable in a small sample size. This study addresses this issue by using Moran’ Index. The variable is considered as global one if Moran’ Index indicates random, while the variable is viewed as local one if it’s not a random distribution. This method may be not that reasonable, but it can also provide a reference for selecting local/global variables to avoid becoming into data-driving.

Back to the position of this issue, comparing with the traditional four-step model, direct station-level transit ridership forecasting model showed its advantages of rapid response, low cost, and efficiency. But on the other hand, the direct model was still a part of the four-step model, which could be viewed as the in-depth first step (forecasting of traffic generation) of the four-step model. With the enrichment and diversification of data, the influence of environmental changes in catchment area on transit ridership can be mastered more accurately with the help of GIS technology.

# Reference

[1] Takashi Nakamura, “A Study on the Relationship between Land Use around Railway Stations and the Railway Station Passengers,” *J. City Plan. Inst. Japan*, vol. 50, no. 3, pp. 1324–1329, 2015.

[2] H. J. Miller, “Potential contributions of spatial analysis to geographic information systems for transportation (GIS-T),” *Geogr. Anal.*, vol. 31, no. 4, pp. 373–399, 1999.

[3] D. E. Boyce, Y.-F. Zhang, and M. R. Lupa, “Introducing ‘ Feedback ’ into Four-Step Travel Forecasting Procedure Versus Equilibrium Solution of Combined Model,” *Transp. Res. Rec.*, vol. 10, no. 12, pp. 123–142, 1994.

[4] X. Chu, “Ridership models at the stop level,” 2004.

[5] R. Cervero, “Alternative Approaches to Modeling the Travel-Demand Impacts of Smart Growth,” *J. Am. Plan. Assoc.*, vol. 72, no. 3, pp. 285–295, 2006.

[6] N. Duduta, “Direct Ridership Models of Bus Rapid Transit and Metro Systems in Mexico City, Mexico,” in *Transportation Research Record: Journal of the Transportation Research Board*, 2013, pp. 93–99.

[7] I. S. Jones and A. J. Nichols, “The demand for inter-city rail travel in the United Kingdom: some evidence,” *J. Transp. Econ. policy*, pp. 133–153, 1983.

[8] G. Walters and R. Cervero, “Forecasting transit demand in a fast growing corridor: the direct-ridership model approach,” *Fehrs Peers Assoc.*, 2003.

[9] E. Guerra, R. Cervero, and D. Tischler, “Half-Mile Circle,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2276, no. 2276, pp. 101–109, 2012.

[10] B. W. Alshalalfah and a. S. Shalaby, “Case Study: Relationship of Walk Access Distance to Transit with Service, Travel, and Personal Characteristics,” *J. Urban Plan. Dev.*, vol. 133, no. 2, pp. 114–118, 2007.

[11] M. J. N. Keijer and P. Rietveld, “How do people get to the railway station? The dutch experience,” *Transportation Planning and Technology*, vol. 23, no. 3. pp. 215–235, 2000.

[12] A. T. Murray, R. Davis, R. J. Stimson, and L. Ferreira, “Public Transportation Access,” *Transp. Res. Part D Transp. Environ.*, vol. 3, no. 5, 1998.

[13] S. O’Sullivan and J. Morrall, “Walking Distances to and from Light-Rail Transit Stations,” *Transp. Res. Rec.*, vol. 1538, no. 1, pp. 19–26, 1996.

[14] F. Zhao, L.-F. Chow, M.-T. Li, I. Ubaka, and A. Gan, “Forecasting Transit Walk Accessibility: Regression Model Alternative to Buffer Method,” *Transp. Res. Rec.*, vol. 1835, no. 1, pp. 34–41, 2003.

[15] J. Gutiérrez, O. D. Cardozo, and J. C. García-Palomares, “Transit ridership forecasting at station level: An approach based on distance-decay weighted regression,” *J. Transp. Geogr.*, vol. 19, no. 6, pp. 1081–1092, 2011.

[16] P. Calthorpe, *The next American metropolis: Ecology, community, and the American dream*. Princeton Architectural Press, 1993.

[17] R. Crane and R. Crepeau, “Does neighborhood design influence travel?: A behavioral analysis of travel diary and GIS data,” *Transp. Res. Part D Transp. Environ.*, vol. 3, no. 4, pp. 225–238, 1998.

[18] R. Cervero, “Built environments and mode choice: Toward a normative framework,” *Transportation Research Part D: Transport and Environment*, vol. 7, no. 4. pp. 265–284, 2002.

[19] B. P. Y. Loo, C. Chen, and E. T. H. Chan, “Rail-based transit-oriented development: Lessons from New York City and Hong Kong,” *Landscape and Urban Planning*, vol. 97, no. 3. pp. 202–212, 2010.

[20] R. Cervero and K. Kockelman, “Travel demand and the 3Ds: Density, diversity, and design,” *Transp. Res. Part D Transp. Environ.*, vol. 2, no. 3, pp. 199–219, 1997.

[21] E. Beimborn, M. Greenwald, and X. Jin, “Accessibility, Connectivity, and Captivity: Impacts on Transit Choice,” *Transp. Res. Rec.*, vol. 1835, no. 1, pp. 1–9, 2003.

[22] J. C. García-Palomares, J. Gutiérrez, and O. D. Cardozo, “Walking accessibility to public transport: An analysis based on microdata and GIS,” *Environment and Planning B: Planning and Design*, vol. 40, no. 6. pp. 1087–1102, 2013.

[23] K. Sohn and H. Shim, “Factors generating boardings at Metro stations in the Seoul metropolitan area,” *Cities*, vol. 27, no. 5. pp. 358–368, 2010.

[24] O. D. Cardozo, J. C. García-Palomares, and J. Gutiérrez, “Application of geographically weighted regression to the direct forecasting of transit ridership at station-level,” *Appl. Geogr.*, vol. 34, no. 4, pp. 548–558, 2012.

[25] J. Choi, Y. J. Lee, T. Kim, and K. Sohn, “An analysis of Metro ridership at the station-to-station level in Seoul,” *Transportation (Amst).*, vol. 39, no. 3, pp. 705–722, 2012.

[26] M.-J. Jun, K. Choi, J.-E. Jeong, K.-H. Kwon, and H.-J. Kim, “Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul,” *J. Transp. Geogr.*, vol. 48, pp. 30–40, 2015.

[27] H. Sung, K. Choi, S. Lee, and S. Cheon, “Exploring the impacts of land use by service coverage and station-level accessibility on rail transit ridership,” *J. Transp. Geogr.*, vol. 36, pp. 134–140, 2014.

[28] M. Kuby, A. Barranda, and C. Upchurch, “Factors influencing light-rail station boardings in the United States,” *Transp. Res. Part A Policy Pract.*, vol. 38, no. 3, pp. 223–247, 2004.

[29] B. D. Taylor, D. Miller, H. Iseki, and C. Fink, “Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas,” *Transportation Research Part A: Policy and Practice*, vol. 43, no. 1. pp. 60–77, 2009.

[30] J. Zhao, W. Deng, Y. Song, and Y. Zhu, “Analysis of Metro ridership at station level and station-to-station level in Nanjing: An approach based on direct demand models,” *Transportation (Amst).*, vol. 41, no. 1, pp. 133–155, 2014.

[31] F. Zhao, L.-F. Chow, M.-T. Li, and X. Liu, “A Transit Ridership Model Based on Geographically Weighted Regression and Service Quality Variables,” *Lehman Cent. Transp. Res. Florida Int. Univ. Miami, Florida. http//lctr. eng. fiu. edu/re-project-link/finalDO97591\\_BW. pdf (accessed December 12, 2010)*, 2005.

[32] Y.-C. Chiou, R.-C. Jou, and C.-H. Yang, “Factors affecting public transportation usage rate: Geographically weighted regression,” *Transportation Research Part A: Policy and Practice*, vol. 78. pp. 161–177, 2015.

[33] G. Thompson, J. Brown, and T. Bhattacharya, “What Really Matters for Increasing Transit Ridership: Understanding the Determinants of Transit Ridership Demand in Broward County, Florida,” *Urban Studies*, vol. 49, no. 15. pp. 3327–3345, 2012.

[34] A. Chakraborty and S. Mishra, “Land use and transit ridership connections: Implications for state-level planning agencies,” *Land Use Policy*, vol. 30, no. 1. pp. 458–469, 2013.

[35] A. Fotheringham, C. Brunsdon, and M. Charlton, “Geographically weighted regression,” vol. 28, no. 2008, pp. 305–308, 2002.

[36] T. and T. Ministry of Land, Infrastructure, “About the current state of public transportation.”

[37] E. Guerra and R. Cervero, “Is a Half-Mile Circle the Right Standard for TODs?,” *ACCESS Mag.*, 2013.

[38] Tadakatsu Nakamura, “An Empirical Analysis of Situations around Station as a Factor Affecting the Number of Station Users,” *Reg. Policy Res.*, vol. 17, no. 3, pp. 15–26, 2015.

[39] C. R. Bhat and J. Y. Guo, “A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels,” *Transp. Res. Part B Methodol.*, vol. 41, no. 5, pp. 506–526, 2007.

[40] T. Nakaya, M. Charlton, P. Lewis, S. Fotheringham, and C. Brunsdon, “GWR4 User Manual,” no. June 2009, p. 40, 2012.

[41] J. Zhao, W. Deng, Y. Song, and Y. Zhu, “What influences Metro station ridership in China? Insights from Nanjing,” *Cities*, vol. 35. pp. 114–124, 2013.

[42] Y. Song and G. J. Knaap, “Measuring the effects of mixed land uses on housing values,” *Reg. Sci. Urban Econ.*, vol. 34, no. 6, pp. 663–680, 2004.