Yikuan Ye
Dec/4/2023
UP 517 Data Science for Planners
Final Project Report

Exploring the effects of built environment factors on metro ridership in Chicago using multiscale geographically weighted regression analysis

1. Introduction

Metro systems globally play a pivotal role in urban transportation, addressing challenges like urban sprawl, congestion, and environmental impact. This is particularly true in metropolitan areas over the globe, where public transit is essential for daily commuting and urban sustainability. As the third-busiest rapid transit system in the United States, the Chicago "L" was recorded an annual ridership of 104 million in 2022 by the Chicago Transit Authority (CTA) and recent data indicates that ridership across Chicago's transit system continues to increase post-Covid. However, the ridership at stations often shows strong spatial differences, and what factors influence the metro ridership has aroused widespread interest among urban researchers and transportation departments.

Prior research has identified various variables influencing metro ridership. These include external factors such as time and weather, land use and socio-economic characteristics, internal factors like transit system quality, intermodal connections (Zhang et al., 2018; Liu et al., 2019; Dadi et al., 2019), and the location of stations. Such studies highlight the intricate relationship between metro ridership and the surrounding environment. In terms of methodology, many models are developed to understand this relationship, including Ordinary Least Squares (OLS), geographically weighted regression, Poisson models, and multiple logit regression models. It is noteworthy that Multiscale Geographically Weighted Regression (MGWR) has demonstrated its advantages in avoiding model underfitting due to spatial heterogeneity in several recent studies (Mansour et al.,2021; Gu et al., 2021; An et al., 2022). It adopts adaptive bandwidths to accommodate the model to spatial non-stationarity and the multiscale effects of each variable, which also have shown its effectiveness in examining the factors influencing metro ridership in Shanghai (Li et al., 2023).

Despite the comprehensive nature of previous studies, there remain gaps in understanding the impacts of built environment on metro ridership in different urban contexts, especially when there is a need to determine the catchment area of metro stations. The catchment area is typically defined as the area around a transit station from which most of its passengers originate, the scale of which varies greatly in different places. This emphasizes the necessity of focused research on how factors within the catchment areas of Chicago's metro stations influence ridership patterns. Thus, the objective of this study is twofold: (1) discovering the specific scale of metro station catchment area (MSCA) in Chicago by employing and comparing different MGWR models based on different scenarios of MSCA and (2) understanding metro ridership spatial variation by taking multiple built environment factors into consideration. These findings can serve as a supplement to other related studies, as well as offer practical insights for local authorities, guiding them in implementing tailored policies that enhance the effectiveness and utilization of metro systems.

2. Data and Methodology

2.1 Study area

To address the transportation needs of a rapidly growing population and to alleviate traffic congestion on the streets, the construction of Chicago "L" was started in the late 19th century. As of 2023, the system includes eight different routes: Red, Blue, Brown, Green, Orange, Purple, Pink and Yellow, and covers a total route length of approximately 102.8 miles (165.4 km). These lines collectively serve 145 stations across the city and its suburbs. In 2022, the CTA recorded an annual ridership of 243.5 million, with the "L" accounting for a significant portion of 42.7%. Obviously, “L” has become a popular choice for city commuters.

This study involves 7 metro lines (Red, Blue, Brown, Green, Orange, Purple, and Pink) connected by a total of 116 stations (Fig.1), which are all within the boundaries of the city of Chicago while the stations in O’Hare International Airport was excluded for its unique built environment. The shapefiles of “L” lines and stations were collected from Chicago City Portal.

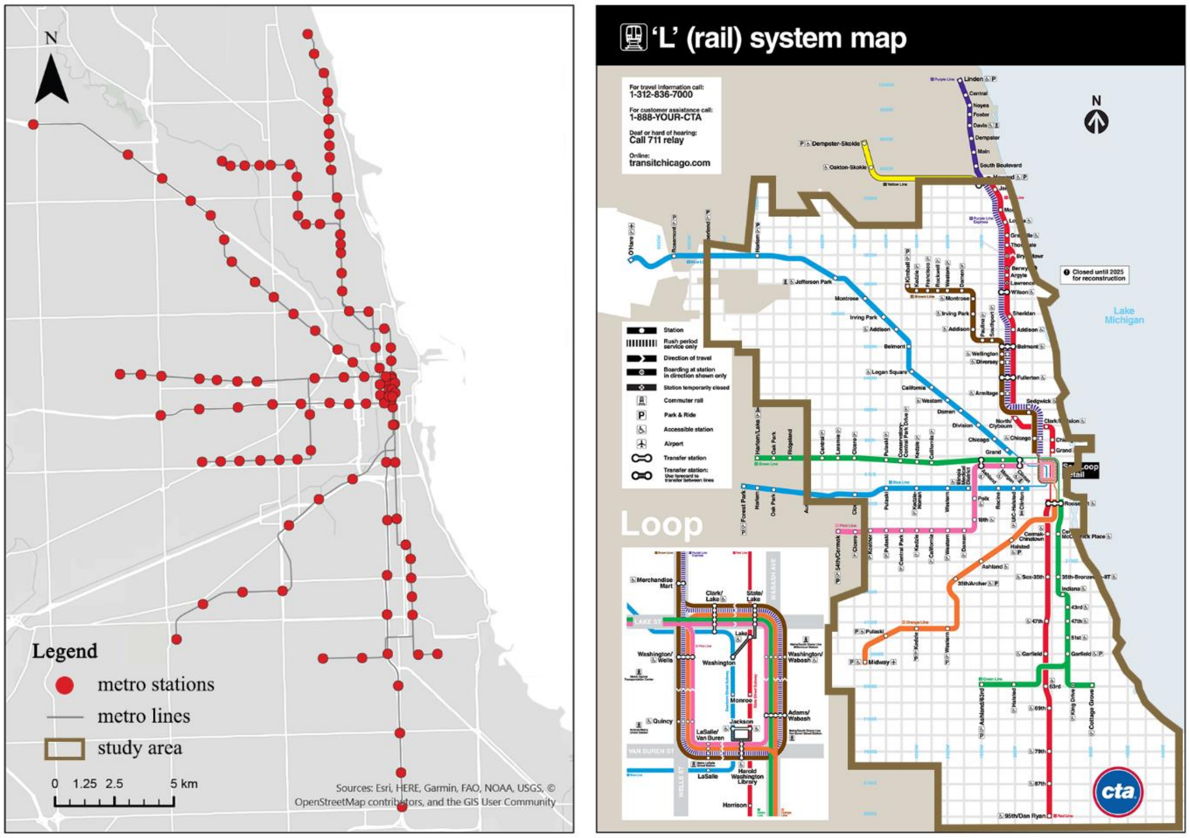

Fig.1 The study area and the "L" system in Chicago (Source: https://www.transitchicago.com).

2.2 Dependent variables

The metro ridership data used in this study was collected from Chicago City Portal and includes four information items: station ID, line name, daily ridership, and day type (weekday, Saturday, Sunday/holidays). To minimize the impact of the Covid-19 pandemic on metro ridership and ensure the data is up-to-date, daily ridership data of each station was used from July 1, 2022, to June 30, 2023, covering a whole year period.

It takes several steps to get the dependent variables ready for model building. First, due to the varied activities that residents engage in during weekdays, weekends, and holidays, there is a significant fluctuation in metro ridership (Fig.2). On weekends and holidays, the metro sees similar levels of daily ridership, both of which are substantially lower compared to weekdays. Hence, I categorized the ridership data into two distinct groups to manage temporal influences: one for weekdays (255 days), and the other for non-weekdays (110 days). The daily ridership on weekdays and non-weekdays were respectively used as two dependent variables.

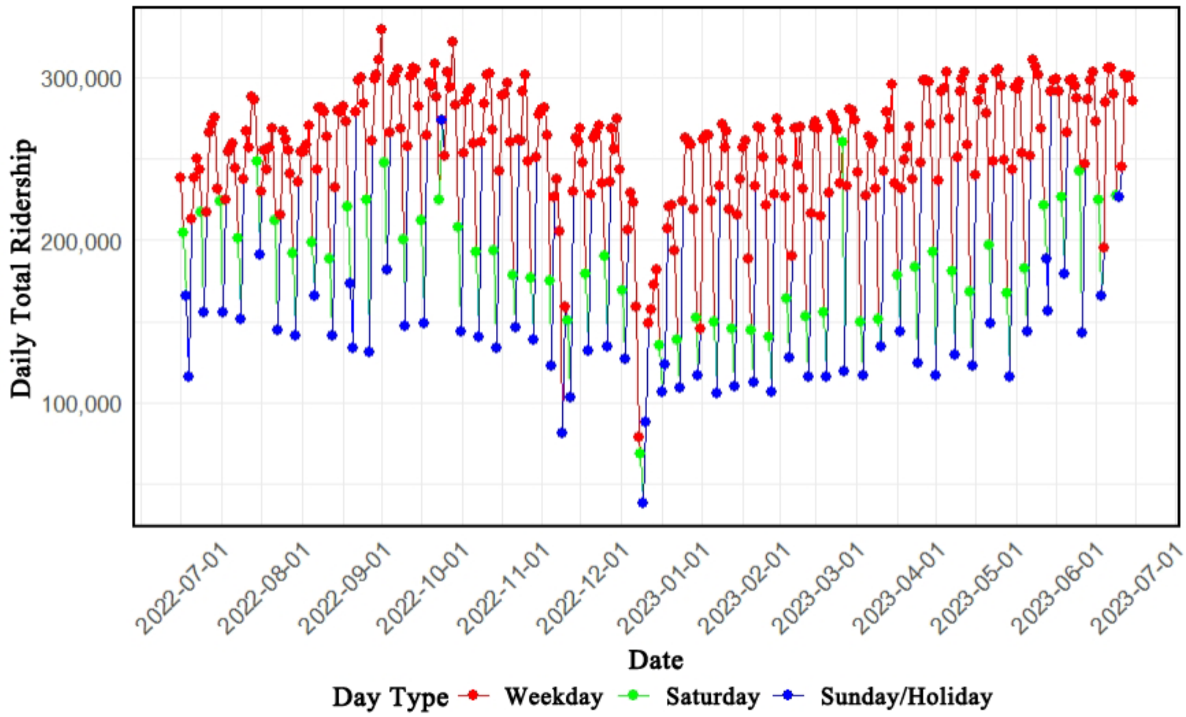

Fig.2 Daily total ridership.

Second, when conducting normality tests on dependent variables, the ridership data exhibits a right/positive skew (Fig.3), which means the bulk of the data is concentrated in the area of smaller values. To align with the presumed Gaussian distribution, a log-normal transformation was performed to the raw data. Thus, the ultimate dependent variable is the log-transformed value of metro ridership.

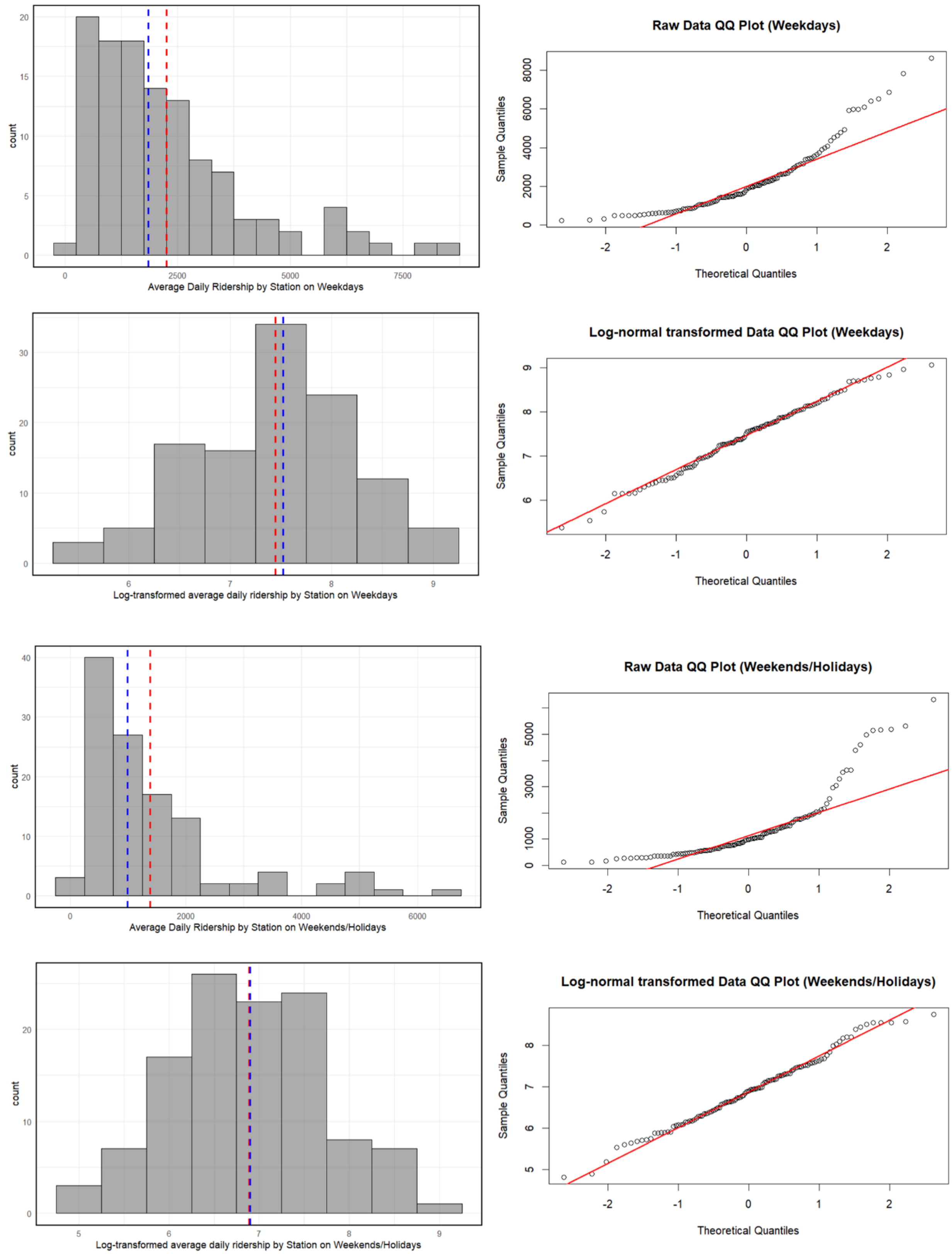

Fig.3 Normality tests on the dependent variables.

2.3 Independent variables

This study aims to uncover the spatial discrepancies in metro ridership caused by built environment factors. Focusing on transportation features (including the number of transfer lines, the number of bus stops and road density) and building and the environment features (including building density and land use), I selected 10 independent variables that may have impacts on the daily metro ridership as follows:

(1) Transfer lines (TL) counts the number of interchange lines available at each station.

(2) Bus stops (BS) is the number of bus stops in the MSCA of each station.

(3) Road density (RD) is the total road length in the MSCA of each station divided by the area of the MSCA.

(4) Building density (BD) is the total building floor area in the MSCA of each station divided by the area of the MSCA.

(5) Land use intensity (LUI) is the number of all POIs in the MSCA of each station.

(6) Land use mixture (LUM) is a variable trying to measure the diversity of POIs in the MSCA of each station, which is calculated in the form of Shannon entropy:

$$LUI = -\sum_{i=1}^{k} p(x_i) \log_k p(x_i)$$

where $P(x_i)$ is the percentage of the i-th type of POIs, k is the total number of POI types.

(7) – (10) Living services (LS), government and business services (GBS), tourism services (TS) and education services (ES) are defined as the percentage of each corresponding type of POIs in MSCA of each station.

These independent variables were derived from the data collected from three sources: Chicago Data Portal, OpenStreetMap and Google Maps Platform. Detailed information is displayed in Table 1.

Table 1

Description and data sources of independent and dependent variables.

Variable	Description	Raw Data (Source)	Abbr.
Dependent variable			
Daily metro ridership by station on weekdays	Log-normal transformed ridership data at each station averaged by weekdays in a year	Daily ridership data of L stations during 7/1/2022-6/30/2023 (Chicago Data Portal)	WD_daily
Daily metro ridership by station on weekends/holidays	Log-normal transformed ridership data at each station averaged by weekends/holidays in a year	Daily ridership data of L stations during 7/1/2022-6/30/2023 (Chicago Data Portal)	NWD_daily
Independent variable			
Transportation features			
Transfer lines	Number of transfer lines at each metro station	Transfer lines at each L station (Chicago Data Portal)	TL
Bus stops	Number of bus stops in the MSCA of each station	Bus stop shapefile in the city of Chicago (Chicago Data Portal)	BS
Road density	Total road length in the MSCA of each station divided by the area of the MSCA.	Road center line shapefile in the city of Chicago (Open Street Map)	RD
Building and the environment features			
Building density	Total building floor area in the MSCA of each station divided by the area of the MSCA	Building footprint shapefile in the city of Chicago (Chicago Data Portal)	BD
Land use intensity	Number of all POIs in the MSCA of each station	POIs data (Google Maps Platform API)	LUI
Land use mixture	Diversity of POIs in the MSCA of each station calculated in the form of Shannon entropy	POIs data (Google Maps Platform API)	LUM
Living services	Percentage of POIs categorized into the living services (catering, shops, sports, parks) in the MSCA of each station	POIs data (Google Maps Platform API)	LS
Government and business services	Percentage of POIs categorized into the working services (public services, companies, industries) in the MSCA of each station	POIs data (Google Maps Platform API)	GBS
Toursim services	Percentage of POIs categorized into the tourism services (attractions, hotels, motels) in the MSCA of each station	POIs data (Google Maps Platform API)	TS
Education services	Percentage of POIs categorized into the education services (educational institutions at all levels) in the MSCA of each station	POIs data (Google Maps Platform API)	ES

2.4 Multiscale geographically weighted regression (MGWR)

Geographically Weighted Regression (GWR) is widely used as a local regression technique which estimate coefficients by using nearby sampling points around a regression point. In traditional GWR, a uniform bandwidth is used for all independent variables that may result in selecting an excessive number of sample points in areas with high data density, while too few points are chosen in sparsely populated areas. This imbalance can reduce the model's accuracy. The basic formula for GWR is as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{K} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i$$

where y_i is the dependent variable, $\beta_0(u_i, v_i)$ is the intercept, $\beta_k(u_i, v_i)$ are local coefficients for each explanatory variable x_{ik}, and ε_i is the error term.

Multiscale Geographically Weighted Regression (MGWR), on the other hand, allows for varying spatial relationships at different scales, which is crucial in handling spatial heterogeneity more effectively. It improves upon the GWR by allowing each variable in the model to have its own spatial scale or bandwidth. This flexibility leads to more precise and localized modeling of spatial data. The formula for MGWR is an extension of the standard GWR model which can be expressed as:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{K} \beta_k(u_i, v_i; h_k)x_{ik} + \varepsilon_i$$

where $\beta_k(u_i,v_i;h_k)$ are the local coefficients for each explanatory variable x_{ik}, where each β_k has its own bandwidth h_k .

In this study, the MGWR 2.2 software developed by Arizonia State University was employed to conduct the MGWR modeling. Additionally, all variables including dependent and independent variables were standardized before modeling to ensure all variables were on a comparable scale.

2.5 Definition of metro station catchment areas (MSCA)

There is no consensus regarding the scale of catchment areas (also referred to as service areas) in the academic field, and practices in transportation planning also vary across different regions. For instance, a study on bus stop ridership in Montreal used different circular buffers with a 200m, 400m, and 800m radius to measure independent variables (Chakour et al., 2016). In metropolises like Shanghai and Beijing in China, 600m and 800m are commonly adopted as the service radius of a metro station. While in the U.S., urban planners have the assumption that the typical person is willing to walk for approximately 10 minutes to reach a transit station, equating to about half a mile measured in a straight line.

In the study, the metro station catchment areas are defined in three different methods:

(1) Circular buffers with a 400m, 500m, 600m and 800m radius are considered, as shown in Fig 4 (a).

(2) Isochrones of 5min and 10min by walking are considered, as shown in Fig 4 (b).

(3) Circular buffers or isochrones overlapped with Thiessen polygons for each station are considered. To reduce computational effort, I limit the overlapping of Thiessen polygons to only those buffers or isochrones that demonstrate better model fitting.

There are 9 independent variables (BS, RD, BD, LUI, LUM, LS, GBS, TS and ES) measured based on the MSCA. With the definition of the MSCA varying, the values measured for variables mentioned above can exhibit significant variation. In this process, multicollinearity tests between

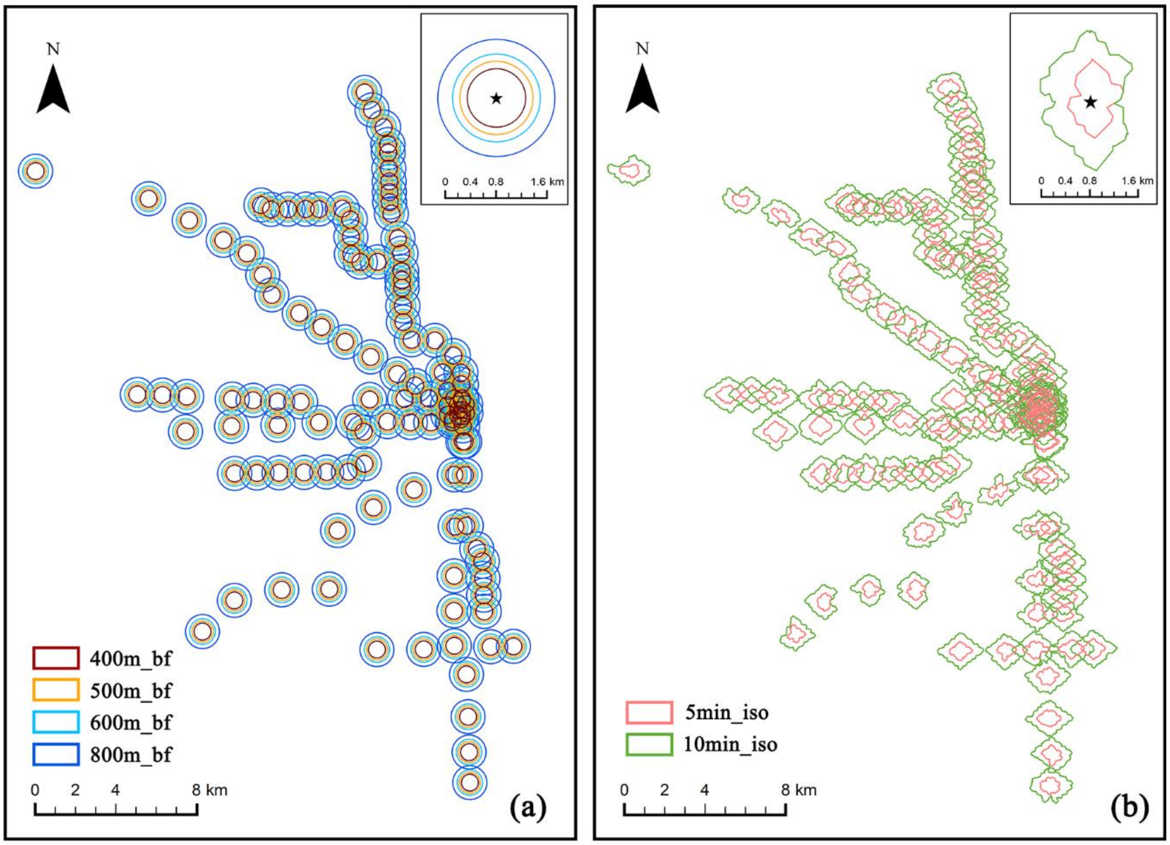

Fig.4 Circular buffers and isochrones for MSCA.

variables are introduced to avoid bias in interpreting the influence of independent variables. To evaluate multicollinearity, a three-step process was followed. Initially, Pearson correlation coefficients (PCCs) were determined for each pair of variables, and those with coefficients exceeding 0.7 were considered to be removed. Next, the variance inflation factors (VIFs) for each variable were calculated, and any variable with VIF over 10 was removed. Finally, condition numbers (CN) were calculated to assess the stability of the regression models, and CN values below 10 would indicate that there is no significant multicollinearity concern in the model. The study has discovered as the coverage area of the MSCA increases, the overlap between them would become more extensive, resulting in the removal of more independent variables due to multicollinearity, as shown in Table 2. When the MSCA were defined as 400m buffers or 5min isochrones by walking, LS and BD were removed. When the MSCA were defined as 500m, 600m buffers or 10min isochrones by walking, LS, BD, BS were excluded. When the MSCA reached 800m buffers, the overlapping was so severe that LS, BD, BS and LUI had to be discarded.

Table 2

Multicollinearity tests between variables measured in 6 scenarios of MSCA
(WD_daily and NWD_daily as the dependent variable have similar results, only the former are shown in the table).

Multicollinearity test method	400m_bf	500m_bf	600m_bf	800m_bf	5min_iso	10min_iso
Pearson Correlation Coefficients (*Remove variables with PCCs > 0.7*)	BD - LUI 0.896 LS - LUM -0.911 ***(LS, BD are removed)***	BS - BD 0.779 BS - LUI 0.786 BD - LUI 0.926 LS - LUM -0.916 ***(LS, BD, BS are removed)***	BS - BD 0.818 BS - LUI 0.837 BD - LUI 0.942 LS - LUM -0.918 ***(LS, BD, BS are removed)***	BS - BD 0.844 BS - LUI 0.854 BS - TS 0.710 BD - RD 0.714 LUI - RD 0.722 BD - LUI 0.961 BD - TS 0.715 LUI - TS 0.700 LS - LUM -0.938 ***(LS, BD, BS, LUI are removed)***	BD - LUI 0.883 LS - ES -0.729 ***(LS, BD are removed)***	BS - BD 0.807 BS - LUI 0.832 BD - LUI 0.946 LS - LUM -0.940 ***(LS, BD, BS are removed)***
Variance Inflation Factors (*Remove variables with VIFs > 10*)	TL 1.507 BS 1.966 RD 1.337 LUI 2.887 LUM 4.975 GBS 2.354 TS 2.290 ES 2.649	TL 1.509 RD 1.474 LUI 2.653 LUM 5.817 GBS 2.389 TS 2.920 ES 3.520	TL 1.515 RD 1.712 LUI 3.197 LUM 6.805 GBS 2.651 TS 3.491 ES 4.172	TL 1.339 RD 1.496 LUM 7.524 GBS 2.146 TS 3.632 ES 4.941	TL 1.460 BS 1.803 RD 1.354 LUI 2.882 LUM 2.184 GBS 1.767 TS 1.975 ES 1.283	TL 1.494 RD 1.876 LUI 3.552 LUM 7.835 GBS 2.615 TS 3.992 ES 4.669
Condition Number (*Check if CN of the model < 10*)	6.501	6.772	8.815	9.581	4.284	8.879

3. Results and discussion

3.1 Comparison of model fitting using different MSCA

Two groups of models, a total of 12 MGWR models, are established with the daily ridership on weekdays and non-weekdays respectively used as the dependent variable while the independent variables are measured based on the 6 scenarios of MSCA mentioned above.

Model performance, indicated by goodness-of-fit (R^2) and AICc values, of 12 MGWR models are displayed in Table 3, which reveals that as the radius of the circular buffer increases, the model's R^2 shows a downward trend while the AICc value increases. In comparison, models based on smaller-radius circular buffers and isochrones have better fitting. When the dependent variable is the ridership on weekdays, the model achieves the best fit when the independent variables are measured based on a 10-minute walking isochrone. When the dependent variable is the ridership on non-weekdays, the best model fit is obtained when the MSCA is reduced to a 400-meter buffer.

Table 3

The R^2 (AICc values) in 6 scenarios of MSCA
(The highest R^2 and lowest AICc values are displayed in bold).

Dependent variable	400m_bf	500m_bf	600m_bf	800m_bf	5min_iso	10min_iso
WD_daily	0.751	0.743	0.709	0.716	0.735	**0.78**
(*Group 1*)	(212.954)	(209.122)	(220.120)	(221.743)	(211.947)	**(196.124)**
NWD_daily	**0.694**	0.657	0.605	0.63	0.655	0.656
(*Group 2*)	(242.177)	(241.488)	(253.169)	(248.740)	**(240.906)**	(244.726)

Based on the comparison of model fitting results, 2 scenarios of MSCA, 10min_iso and 400m_bf, were selected to further get overlapped with Thiessen polygons (Fig.5). The reason why using Thiessen polygons is that they can divide an independent and non-overlapping service area for each metro station which may help increase stability of the model.

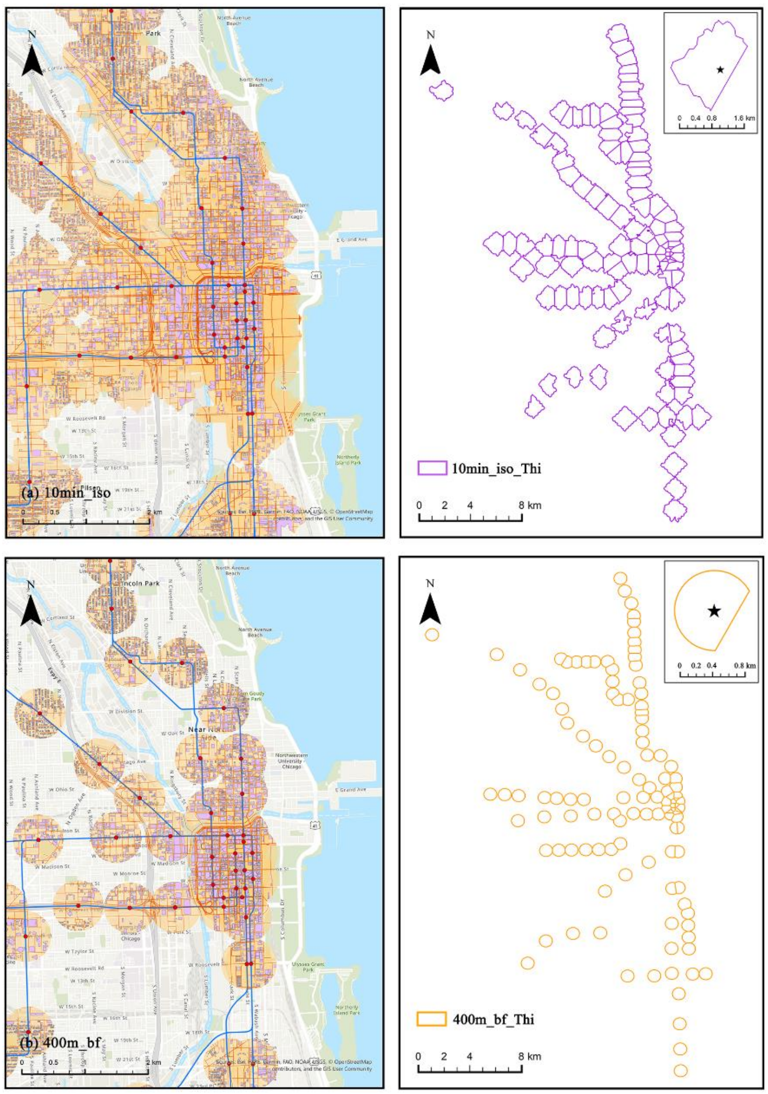

Fig.5 400m circular buffers and 10min walking isochrones overlapped with Thiessen polygons.

The multicollinearity tests with PCCs, VIFs and CN were also conducted on these two new scenarios, which only made LS removed from independent variables. As presented in Table 4, the R^2 and AICc values suggest that the fit of the models not only failed to improve, but actually worsened, surprisingly.

Table 4

The R^2 (AICc values) in 2 better model-fit scenarios of MSCA overlapped with Thiessen polygons (The highest R^2 and lowest AICc values are displayed in bold).

Dependent variable	400m_bf_Thi	10min_iso_Thi
WD_daily	0.705	**0.715**
(*Group 1*)	(234.643)	**(230.761)**
NWD_daily	**0.64**	0.606
(*Group 2*)	**(257.568)**	(262.383)

I believe the reason for this situation is that although the purpose of Thiessen polygons is to reduce the overlap of independent variables between stations, this approach might lead to inaccuracies in reflecting the actual conditions of each station's catchment area. This issue becomes particularly pronounced in densely packed areas like the Loop in downtown Chicago, where minor changes in independent variables can significantly affect model outcomes. The use of Thiessen polygons has resulted in excessively small service areas for some stations, potentially removing built environment factors that simultaneously affect multiple stations, thereby leading to a distortion of the independent variables.

The R^2 and AICc values of the 16 MGWR models are displayed in Fig. 6, clearly showing that the best model fits are under the scenarios of 400m_bf and 10min_iso.

This addresses the first question posed by the study. The specific scale of the catchment area for metro stations in Chicago is as follows: during weekdays, it corresponds to the range of 10-minute walking, while on weekends and holidays, it slightly decreases to a radius of 400 meters, as shown in Fig.7.

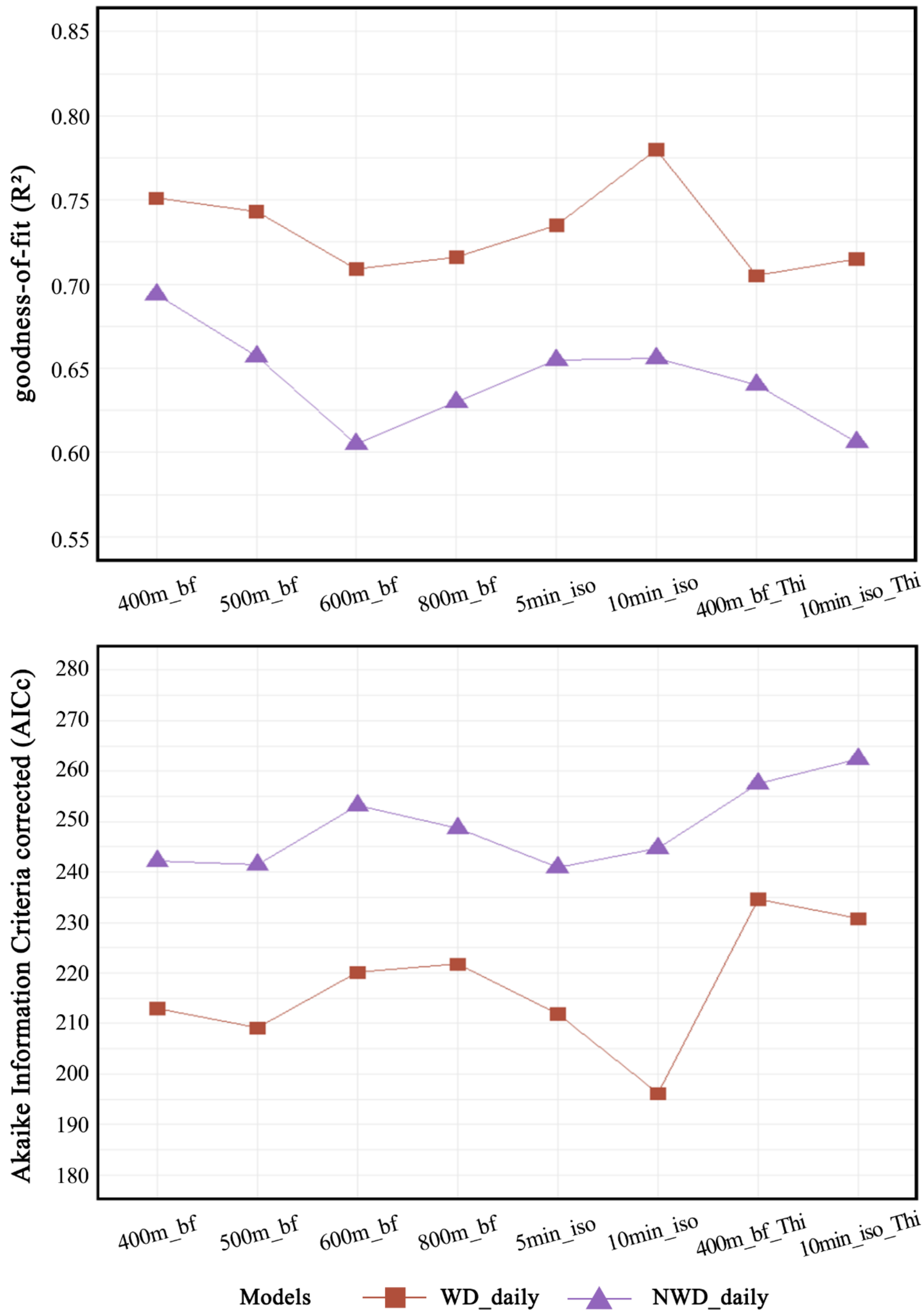

Fig.6 The R^2 and AICc values of 16 MGWR models using independent variables based on different MSCA.

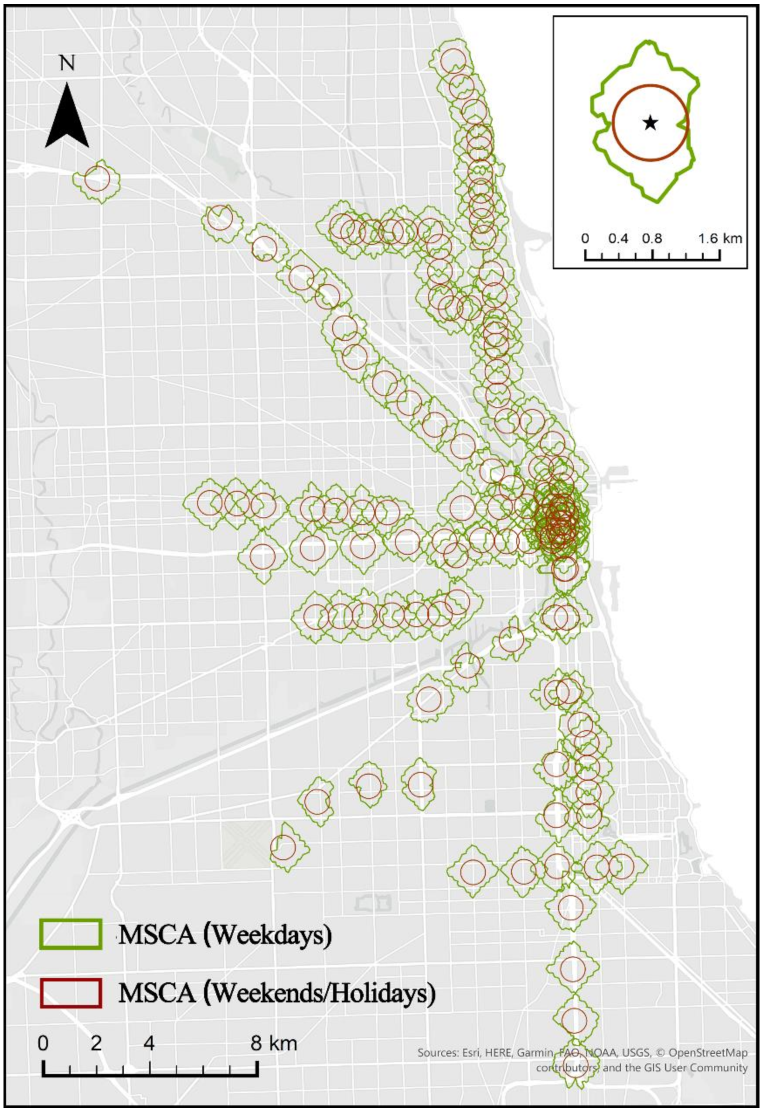

Fig.7 The catchment area for metro stations in Chicago.

The spatial distribution of local R^2 and the residuals of the two models are displayed in Fig.8, respectively. We can see that the local R^2 values for both models are primarily concentrated between 0.65 and 0.85, indicating they have a relatively good fit, while WD_daily (Model 1) has more high values in local R^2, showing its model fit is better than NWD_daily (Model 2). In addition, their distributions are fairly uniform, suggesting stability of the two models. As for residuals, most of the residual values for both models are quite small, signifying low prediction errors. Additionally, the residuals appear to be randomly distributed without any obvious patterns or trends, indicating a good fit of the models.

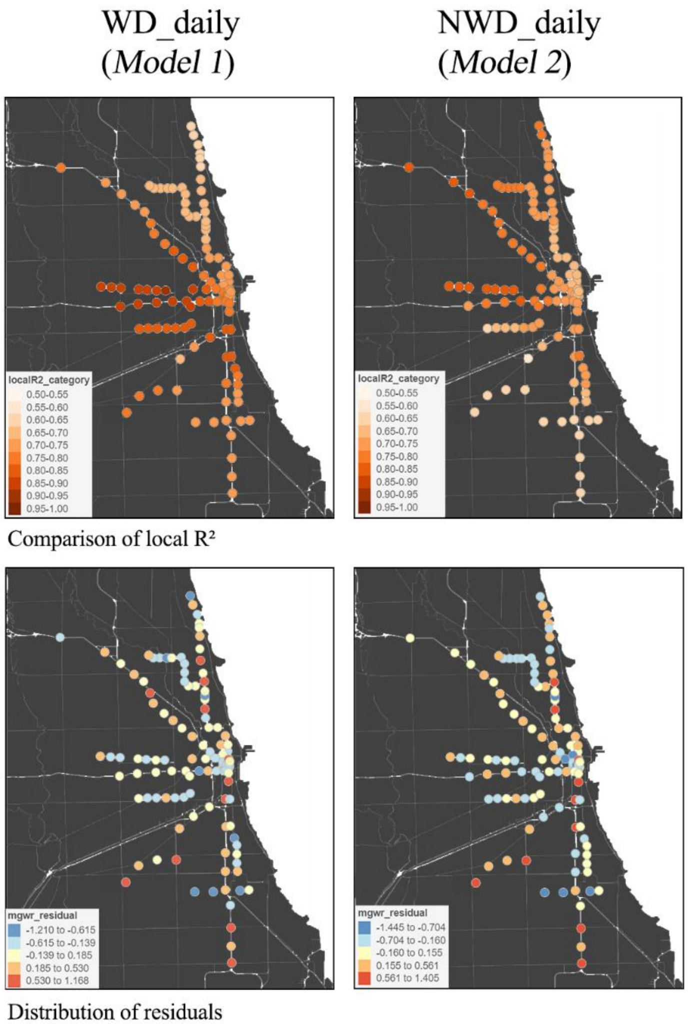

Fig.8 Spatial distribution of local R^2 and residuals of the two models.

3.2 Spatial analysis of influential factors on metro ridership

3.2.1 Spatial patterns of explanatory variables

In the MSCA under the two scenarios of 10min_iso and 400m_bf, the spatial distribution of the independent variables is shown in Fig.9 and Fig.10, respectively. From these maps, some patterns can be observed. Specifically, TL and BS are more densely distributed in the city center area, which may indicate higher transport accessibility and larger ridership, while also imply a potential over-concentration and could lead to relatively insufficient services in the surrounding areas. RD is higher in certain areas, which may correspond to higher ridership and better metro station accessibility. High road density could mean traffic congestion, which might also lead to some passengers turning to use the metro. High values of LUI are concentrated in the city center, while higher LUM is found in the western and southern parts of the city. Besides, GBS is also more concentrated in the city center, while TS and ES are more densely distributed in certain specific areas. These locations will act as key areas for attracting passengers.

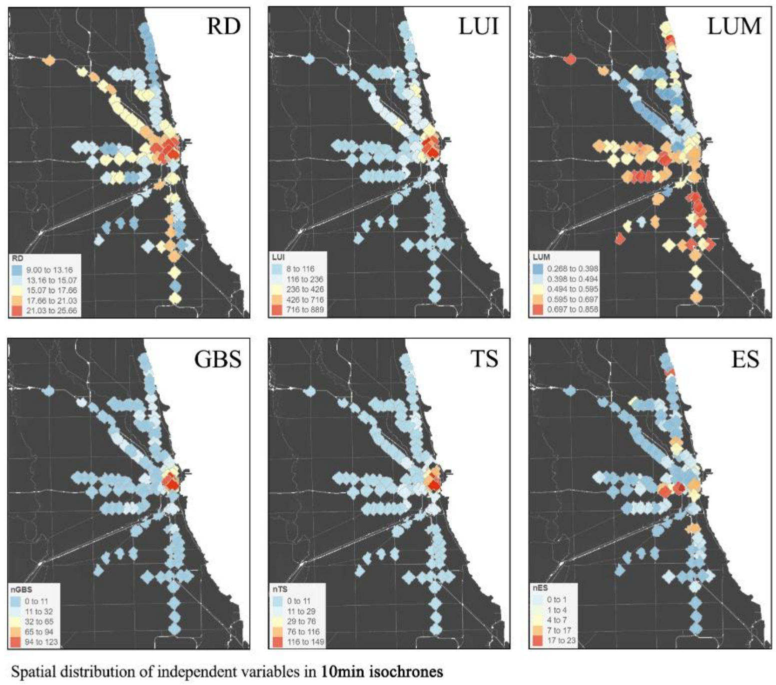

Fig.9 Spatial distribution of independent variables in the MSCA of 10min_iso.

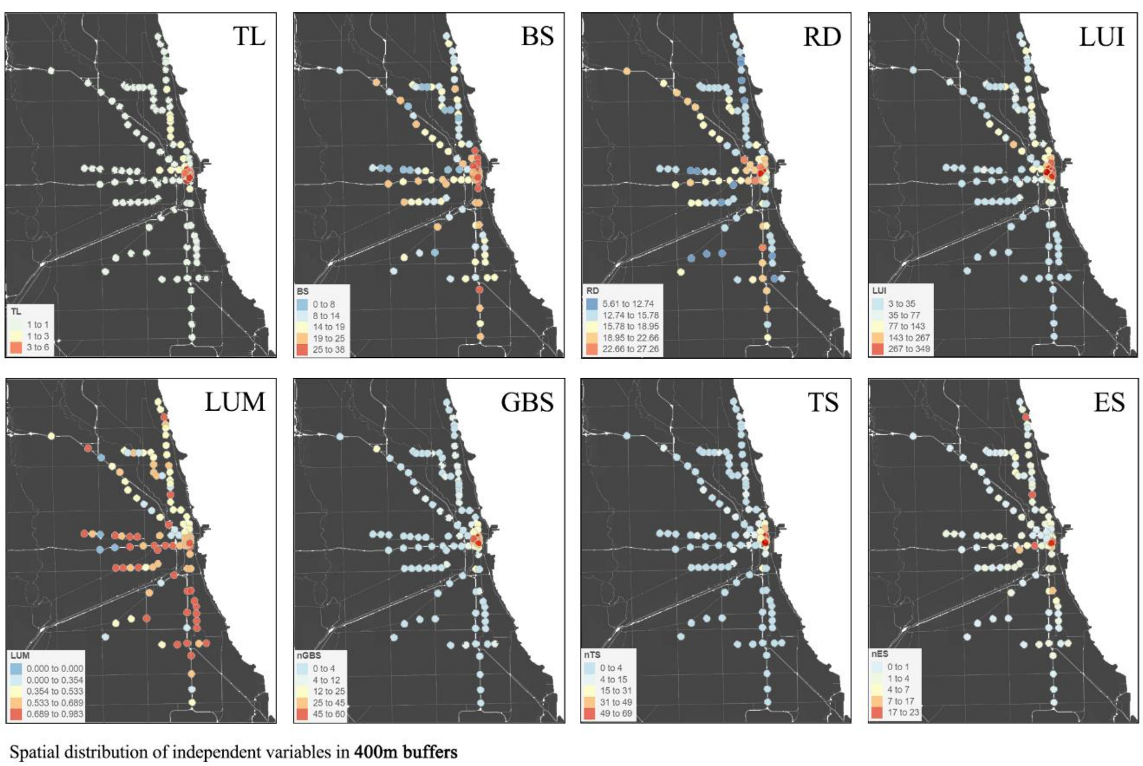

Fig.10 Spatial distribution of independent variables in the MSCA of 400m_bf.

As demonstrated in Section 2.4, MGWR permits each variable to have its own bandwidth. Essentially, this bandwidth is the count of sampling points around a metro station, influencing the determination of the regression coefficient. As illustrated in Fig.11 (a) and (b), the red lines show the variable bandwidths of independent variables, which range from 43-115. The grey bar plot shows the number of significant stations of each variable. A high grey bar indicates the significance of corresponding variable.

From Fig.11 (a) and (b), we can see that there are 4 variables including LUI, TL, GBS and RD have a significant impact on ridership on both weekdays and non-weekdays. They all have a relatively small bandwidth (43-60), indicating their influences are localized. Education services (ES) show significance on weekdays with a large bandwidth (115), indicating its large range of impact on weekday metro ridership which may mean that students and faculty are willing to go long distances to get to and from a metro station.

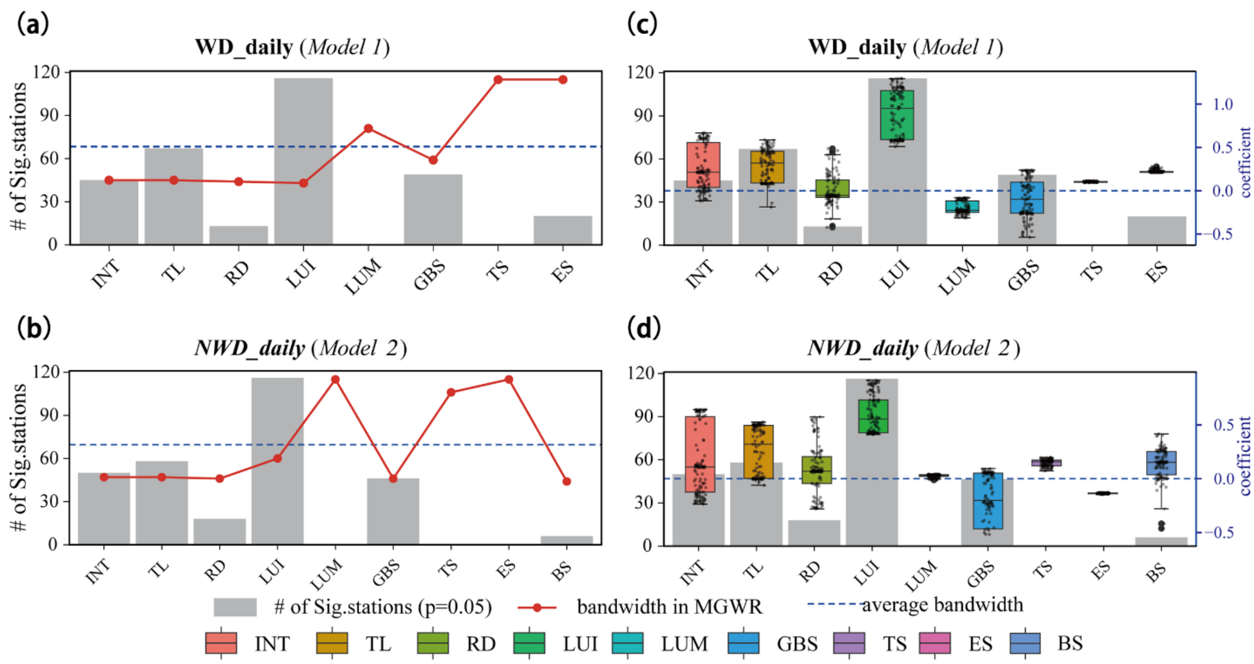

Fig.11 Bandwidths and coefficients of independent variables and the number of significant stations in the two models.

3.2.2 Spatial patterns of regression coefficients

The box plots of the coefficients of independent variables are illustrated in Fig.10 (c) and (d). From these plots, we can identify whether a variable has a positive, negative or both positive and negative impact on the ridership in the two models as well as the magnitude of the impact. It is clear that LUI has the strongest positive impact on ridership on both weekdays and non-weekdays. TL also has positive effect on ridership at the vast majority of stations. On the contrary, GBS has a strong negative impact while RD shows negative impact on weekdays and positive impact on weekends and holidays. Additionally, ES and BS also show a positive impact on metro ridership.

From the spatial patterns of variables' coefficients illustrated in Fig.12, we can further study their impacts and the underlying meanings in a straightforward way. In terms of LUI, we can see a very strong positive impact over all stations, indicating more developed areas with more mixed-use developments contribute to higher ridership. While the impact in the downtown area is relatively small, which may show the development of the city center is nearing saturation, and the increase in land use intensity may not have a significant positive impact on ridership. In terms of TL, at all stations positive impact is shown except for the downtown area. The design of the Loop seems to make little contribution to the ridership. As for GBS, a very strong negative impact is shown at almost all locations, especially for non-weekdays, which means people would prefer to take other means of transportation rather than metro to get to these places. Similarly, the patterns of the coefficients of RD manifest people may prefer to use cars over metro in most places since more road density corresponds to less metro ridership. However, in the south side, RD exert positive effects on metro ridership. Given that high road density leading to high accessibility of metro stations may not fully explain this phenomenon as high road density exerts negative effects on metro ridership in the downtown area, the existence of education institutions in the south side especially the University of Chicago may contribute to this pattern for many students have to take public transportation without cars.

4. Conclusions

After exploring the effects of built environment factors on metro ridership in Chicago by using MGWR models, two major conclusions are drawn as follows:

(1) Independent variables, when assessed using the 10min walking isochrones and 400m buffers as the MSCA, are most effective in explaining metro ridership in Chicago on weekdays and non-weekdays respectively. While it is observed that in areas with closely situated stations and large MSCA, substantial overlap in catchment areas would lead to significant multicollinearity among these variables, Thiessen polygons should be

carefully used since it may fail to reflect the actual situations that there are built environment factors that simultaneously affect multiple stations, especially for areas like downtown Chicago where so many lines are overlapped together and stations are close to each other.

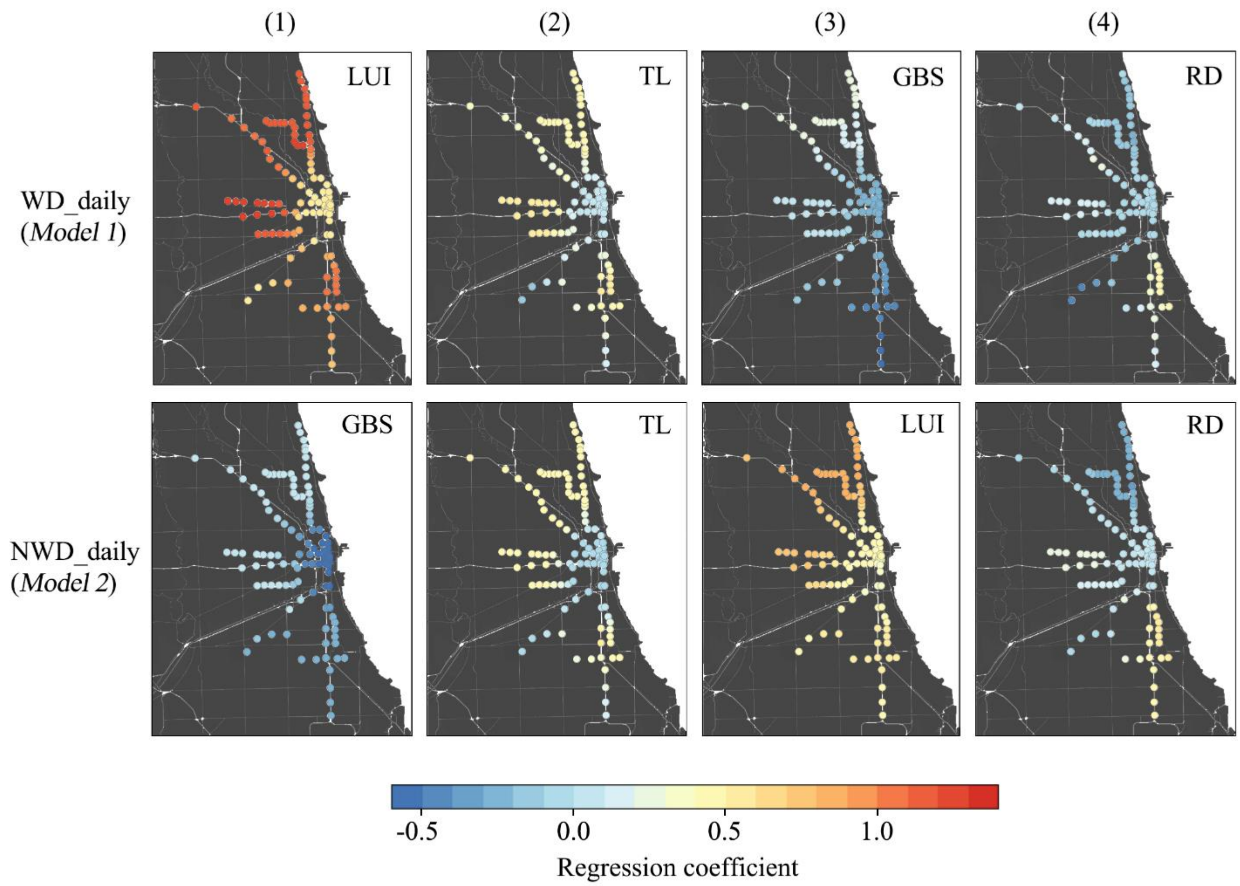

Fig.12 Spatial patterns of coefficients of the 4 variables that have significant impact on metro ridership in the two models.

(2) Spatial patterns from MGWR models indicate that land use intensity (LUI), transfer lines (TL), government and business services (GBS) and road density (RD) have significant impact on ridership and the magnitude of effects varys across different locations and different models. Planning implications from these patterns include the following points: a) Transit-Oriented Development (TOD) should be paid attention to especially in the peripheral areas of Chicago. b) To increase ridership in downtown Chicago, improvements can be made by enhancing the connections between metro stations and other modes of transport like buses or shared bicycles, as well as by improving the

quality and efficiency of the existing metro services rather than simply overlap more metro lines. c) To encourage the use of public transit, business and public services can be planned strategically along the metro lines and the accessibility between facilities and stations should be improved.

Bibliography

A. Stewart Fotheringham, Wenbai Yang & Wei Kang (2017) Multiscale Geographically Weighted Regression (MGWR), Annals of the American Association of Geographers, 107:6, 1247-1265, https://doi.org/10.1080/24694452.2017.1352480.

Dadi An, Xin Tong, Kun Liu, Edwin H.W. Chan, Understanding the impact of built environment on metro ridership using open source in Shanghai, Cities, Volume 93, 2019, Pages 177-187, ISSN 0264-2751, https://doi.org/10.1016/j.cities.2019.05.013.

Fan Gao, Linchuan Yang, Chunyang Han, Jinjun Tang, Zhitao Li, A network-distance-based geographically weighted regression model to examine spatiotemporal effects of station-level built environments on metro ridership, Journal of Transport Geography, Volume 105, 2022, 103472, ISSN 0966-6923, https://doi.org/10.1016/j.jtrangeo.2022.103472.

Gu, H, Yu, H, Sachdeva, M, Liu, Y. Analyzing the distribution of researchers in China: An approach using multiscale geographically weighted regression. Growth and Change. 2021; 52: 443–459. https://doi.org/10.1111/grow.12453.

Mengya Li, Mei-Po Kwan, Wenyan Hu, Rui Li, Jun Wang, Examining the effects of station-level factors on metro ridership using multiscale geographically weighted regression, Journal of Transport Geography, Volume 113, 2023, 103720, ISSN 0966-6923, https://doi.org/10.1016/j.jtrangeo.2023.103720.

Rui An, Zihao Wu, Zhaomin Tong, Sixian Qin, Yi Zhu, Yaolin Liu, How the built environment promotes public transportation in Wuhan: A multiscale geographically weighted regression analysis, Travel Behaviour and Society, Volume 29, 2022, Pages 186-199, ISSN 2214-367X, https://doi.org/10.1016/j.tbs.2022.06.011.

Shawky Mansour, Abdullah Al Kindi, Alkhattab Al-Said, Adham Al-Said, Peter Atkinson, Sociodemographic determinants of COVID-19 incidence rates in Oman: Geospatial modelling

using multiscale geographically weighted regression (MGWR), Sustainable Cities and Society, Volume 65, 2021, 102627, ISSN 2210-6707, https://doi.org/10.1016/j.scs.2020.102627.
Vincent Chakour, Naveen Eluru, Examining the influence of stop level infrastructure and built environment on bus ridership in Montreal, Journal of Transport Geography, Volume 51, 2016, Pages 205-217, ISSN 0966-6923, https://doi.org/10.1016/j.jtrangeo.2016.01.007.
Xiaohu Zhang, Yang Xu, Wei Tu, Carlo Ratti, Do different datasets tell the same story about urban mobility — A comparative study of public transit and taxi usage, Journal of Transport Geography, Volume 70, 2018, Pages 78-90, ISSN 0966-6923, https://doi.org/10.1016/j.jtrangeo.2018.05.002.

Data Sources

City of Chicago | Data Portal | City of Chicago | Data Portal
OpenStreetMap
Google Maps Platform - Location and Mapping Solutions