

Article

Study on Spatially Nonstationary Impact on Catering Distribution: A Multiscale Geographically Weighted Regression Analysis Using POI Data

Lu Tan ^{1,*} and Xiaojun Bu ^{1,2}

¹ College of Forestry and Grassland, Nanjing Forestry University, Nanjing 210037, China

² College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China

* Correspondence: tanlu@njfu.edu.cn

Abstract: Factors related to catering distribution are typically characterized by local changes, but few studies have quantitatively investigated the inherent spatial nonstationarity correlations. In this study, a multiscale geographically weighted regression (MGWR) model was adopted to locally examine the impact of various factors on catering distribution, which were obtained through a novel method incorporating GeoDetector analysis and exploratory factor analysis (EFA) using point of interest (POI) data. GeoDetector analysis was used to identify the effective variables that truly contribute to catering distribution, and EFA was adopted to extract interpretable latent factors based on the underlying structure of the effective variables and thus eliminate multicollinearity. In our case study in Nanjing, China, four primary factors, namely commuting activities, shopping activities, tourism activities, and gathering activities, were retained from eight categories of POIs with respect to catering distribution. The results suggested that GeoDetector working in tandem with EFA could improve the representativeness of factors and infer POI configuration patterns. The MGWR model explained the most variations (adj. R^2 : 0.903) with the lowest AICc compared to the OLS regression model and the geographically weighted regression (GWR) model. Mapping MGWR parameter estimates revealed the spatial variability of relationships between various factors and catering distribution. The findings provide useful insights for guiding catering development and optimizing urban functional spaces.

Keywords: catering distribution; spatial nonstationarity; multiscale geographically weighted regression; exploratory factor analysis; POI data



Academic Editors: Wolfgang Kainz and Wei Huang

Received: 20 December 2024

Revised: 25 February 2025

Accepted: 5 March 2025

Published: 6 March 2025

Citation: Tan, L.; Bu, X. Study on Spatially Nonstationary Impact on Catering Distribution: A Multiscale Geographically Weighted Regression Analysis Using POI Data. *ISPRS Int. J. Geo-Inf.* **2025**, *14*, 119. <https://doi.org/10.3390/ijgi14030119>

Copyright: © 2025 by the authors. Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The catering industry is one of the most prominent sectors of the economy. Despite evolving economic landscapes, catering has demonstrated sustained growth over recent decades, contributing to socioeconomic development worldwide. In China, according to data from the National Bureau of Statistics, the catering industry has been one of the fastest-growing economic sectors in recent years, accounting for nearly 10% of the total retail revenue in 2022. The performance of the catering business is largely dependent on the strategic positioning of catering establishments [1,2], which also serve as fundamental elements in urban space organization [3]. Therefore, investigating catering distribution and its environmental interactions not only empowers the catering business to optimize location strategies, but also assists urban policymakers in enhancing functional space utilization and promoting the development of the catering industry.

The spatial distribution of the catering industry has been extensively investigated following the central place theory [4]. Previous studies indicated that multiple factors may contribute to the spatial distribution of catering in a city, including population density, traffic accessibility, land use, economic conditions, and competition [5–10]. Recent studies have shown that the relationships between catering distribution and its influencing factors exhibit local variations, a phenomenon commonly referred to as spatial nonstationarity. For instance, the impact of population density varies across different catering spatial structures [11], while street centrality also exerts a spatially heterogeneous influence [12]. These findings emphasize the importance of incorporating spatial nonstationarity considerations when investigating the relationships between catering distribution and its influencing factors, particularly in urban-scale analyses [13].

Local regression models, especially the geographically weighted regression (GWR), are commonly used to explore spatially nonstationary characteristics not only in urban geography, but also in other fields [14,15]. Unlike global regression models that employ uniform parameters, GWR enables localized parameter estimation to capture spatial variations [16]. However, GWR has an accuracy problem due to the assumption that all spatially varying processes operate at the same spatial scale. The development of multiscale geographically weighted regression (MGWR) by Fotheringham et al., (2017) addresses this limitation by permitting variable bandwidths for distinct spatial relationships [17]. Empirical studies have demonstrated MGWR's superior performance in handling both spatial nonstationarity and varying spatial scales of regression relationships, thereby achieving more robust significance test results than the GWR and conventional regression models [18]. However, the application of MGWR to catering distribution analysis remains underdeveloped [12].

Previous studies on catering distribution relied heavily on “ready-made” data from government statistics or surveys, which are limited in spatial resolution and scope due to the high cost of surveys [19,20]. With the rapid development of the big data technology, new geographic data sources are constantly being explored, such as point of interest (POI) data. POIs represent locations where people gather and conduct daily activities [21]. POI density can directly or indirectly reflect urban land use and functional zoning [22]. Moreover, POI data based on location-based services have advantages over statistical data and first-hand survey data in terms of both large available amounts and high precision. Therefore, they have been increasingly applied to studies on the urban catering industry [23–25]. However, these studies primarily focus on analyzing the spatial pattern of catering points or clusters, with limited attention given to the relationships between the catering points and the other categories of POIs (i.e., traffic facilities, scenic spots, markets, etc.), which reflect critical determinants of catering distribution, such as traffic accessibility, tourism context, and economic development.

Overall, most studies focus on the spatial distribution pattern of the catering industry within specific cities while neglecting quantitative analyses of the spatially nonstationary relationship between catering distribution and its related factors. The emergence of POI data provides a valuable opportunity for further research in this field. Therefore, in this study, various categories of POIs were collected to represent a set of candidate variables. GeoDetector analysis [26,27] and exploratory factor analysis (EFA) [28] were applied in tandem to extract interpretable factors and solve the problem of multicollinearity. To deal with the abovementioned feature of spatial nonstationarity, the MGWR model, compared with the GWR model and the OLS model, was implemented to analyze the spatially nonstationary impact of factors on catering distribution.

The main contributions of this study are as follows: (1) a novel method incorporating GeoDetector analysis and EFA was employed to identify the factors related to catering dis-

tribution using POI data; (2) the impact of the various factors was examined by quantitative analysis, with particular attention to spatial nonstationarity.

The remainder of the paper is structured as follows. Section 2 introduces the study area and data sources. Section 3 presents the main research methods, including GeoDetector analysis, EFA, and MGWR modeling. Section 4 shows the results obtained from the analysis, along with technical discussions in Section 5. Section 6 concludes the paper.

2. Study Area and Data

2.1. Study Area

Nanjing was adopted as the study area in this study. The city of Nanjing ($31^{\circ}14'–32^{\circ}37' N$, $118^{\circ}22'–119^{\circ}14' E$) is located in the central part of the Yangtze Delta metropolitan area, close to a number of advanced cities in China, such as Shanghai, Hangzhou, and Suzhou. It consists of 11 districts, namely Gulou, Xuanwu, Jianye, Qinhuai, Qixia, Yuhuatai, Pukou, Luhe, Jiangning, Lishui, and Gaochun, with an area of 6587.02 km² (Figure 1). The population reached approximately 9.5 billion in 2022.

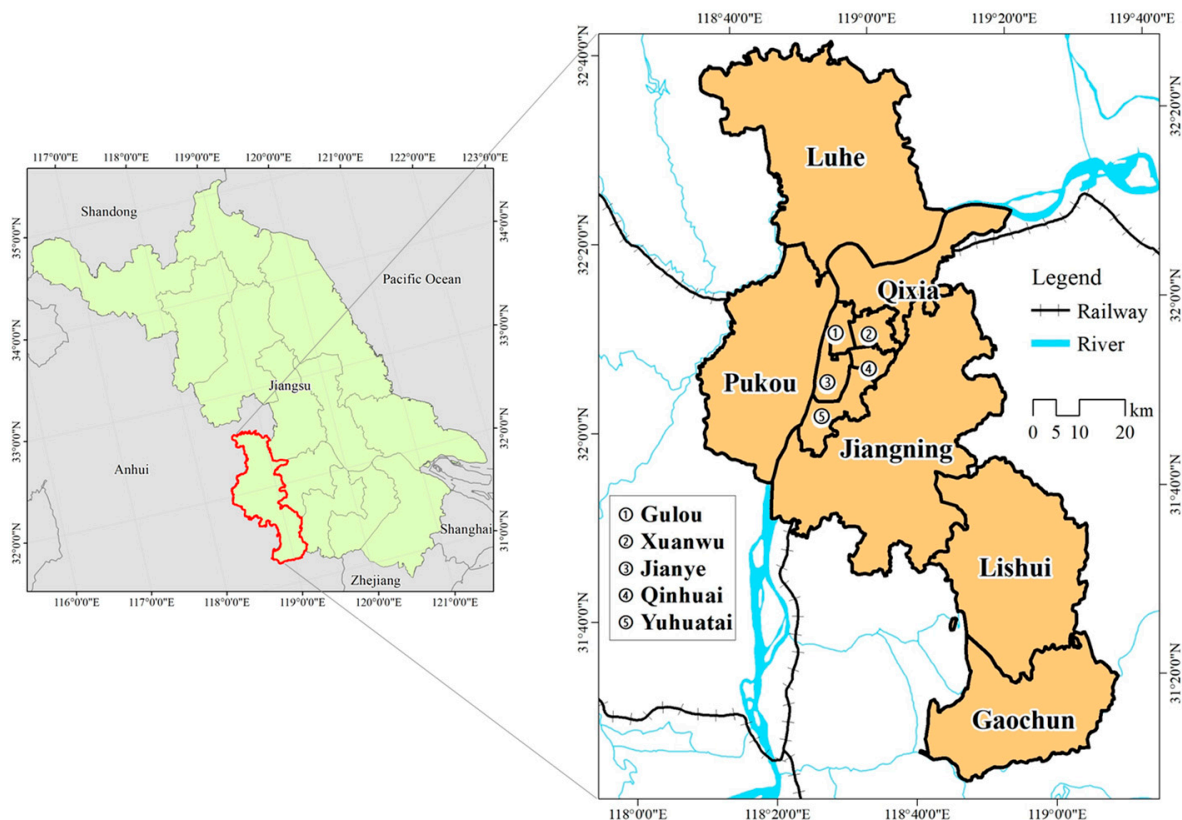


Figure 1. Location of the study area.

As the capital of Jiangsu Province and the central city of the Yangtze Delta metropolitan area, Nanjing is one of the most economically developed cities in China. According to data from the Nanjing Bureau of Statistics, the GDP of the city grew by 55.1 billion yuan in 2022 and reached approximately 1.7 trillion yuan, accounting for 1.35% of the national GDP. Along with rapid economic development, the catering industry of Nanjing has reached a significant level in China in terms of market size, industrial structure, and richness. However, the development of catering space is imbalanced across the city [3]. Therefore, in this study, we took Nanjing as the study area to investigate the spatially nonstationary impact on catering distribution. The findings are expected to provide practical guidance for

optimizing the spatial structure of the catering industry and promoting the development of various types of urban functional spaces in a coordinated manner.

2.2. Data Collection and Preparation

The POI data adopted in this study were obtained from the Baidu Map API (<https://lbsyun.baidu.com>) in August 2023. In total, 26,648 catering points were identified within the study area, including non-fast food restaurants, fast food restaurants, cafes, tea houses, beverage shops, pastry shops (including dessert houses), and snack bars. Eight categories of POIs were collected as candidate variables. Table 1 provides the variable names together with the corresponding types of POIs.

Table 1. Candidate variables used in this study together with descriptions.

Variable (POI)	Description	Unit
Transportation	Transportation facilities, including bus stops, subway stations, parking lots, etc.	count/km ²
Residential	Residential buildings	count/km ²
Office	Enterprises, government agencies, and financial services	count/km ²
Commerce	Shopping malls and integrated shopping centers	count/km ²
Education	Educational institutions, including elementary schools, secondary schools, college campuses, training institutions, etc.	count/km ²
Healthcare	Healthcare services, including hospitals, clinics, drugstores, etc.	count/km ²
Tourism	Scenic spots	count/km ²
Venues	Large venues, including conference centers, exhibition halls, etc.	count/km ²

All the datasets were projected onto the Asia North Albers Equal Area Conic coordinate system. According to the principles of appropriate statistical unit selection [29], a fishnet with a resolution of 2 km × 2 km was used to calculate the density of catering points (dependent variable) and the candidate variables. To ensure data integrity, the units at the edge of the boundary that were not completely within the study area were systematically excluded from the analysis. All the variables were joined to the corresponding statistical units in ArcGIS (v 10.7), totaling 1579 units.

Since the range of variable values varies widely, a min–max normalization method was applied to rescale the data into the range of [0, 1]. The formula is as follows:

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (1)$$

where x^* is the normalized value, x is the original value, x_{max} is the maximum value, and x_{min} is the minimum value. Figure 2 illustrates the spatial distributions of normalized candidate variables and catering point density.

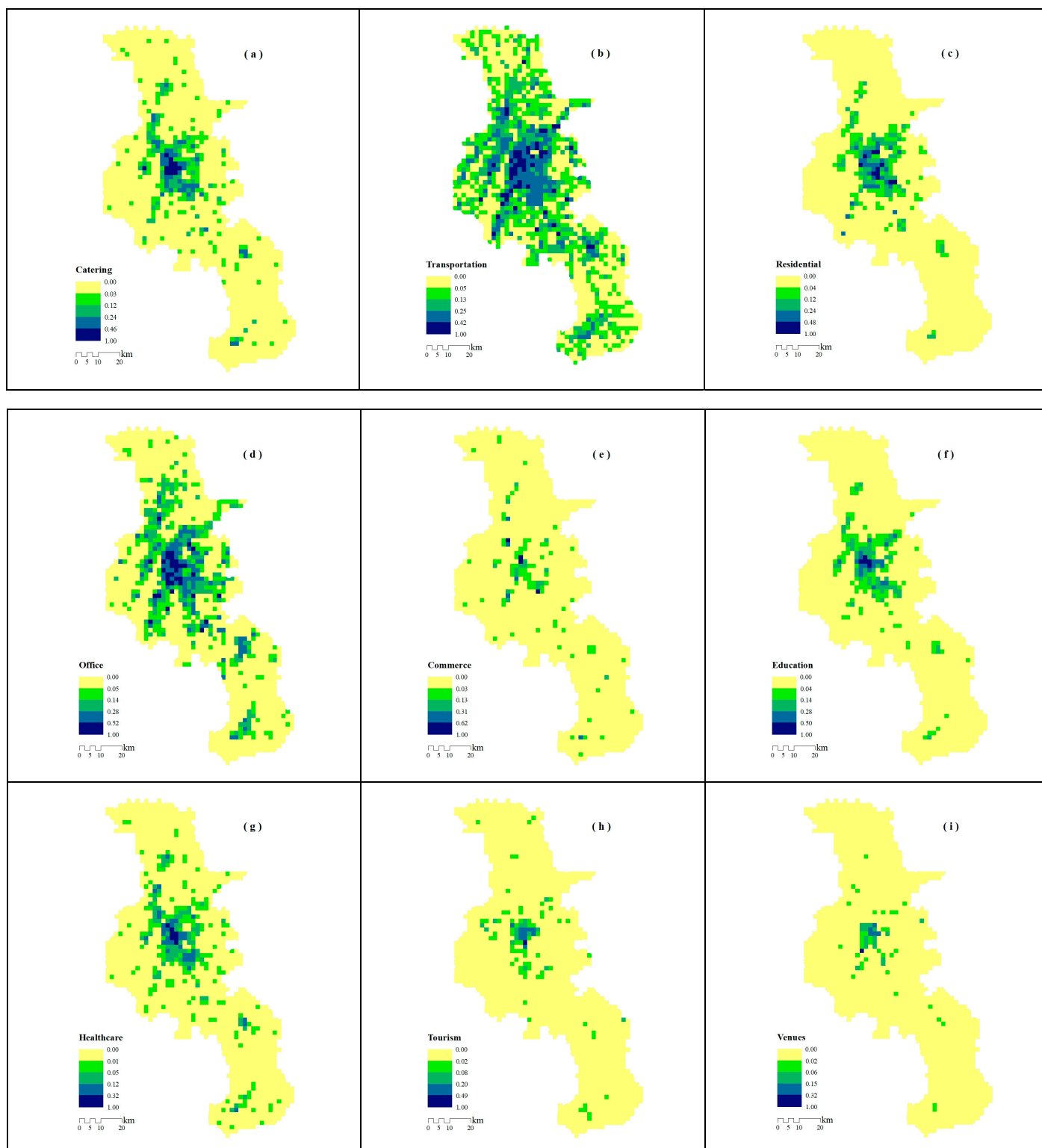


Figure 2. The normalized spatial distribution of (a) catering point density and the candidate variables: (b) transportation, (c) residential, (d) office, (e) commerce, (f) education, (g) healthcare, (h) tourism, and (i) venues.

3. Methodology

3.1. Research Framework

The spatially nonstationary impact on catering distribution was explored through the following steps (Figure 3). First, in order to select the effective variables related to catering

distribution, GeoDetector analysis (detailed in Section 3.2) was performed to compute the contribution of each candidate variable and the synergies between the candidate variables. Thereafter, EFA (detailed in Section 3.3) was conducted to identify several latent factors based on the underlying structure of the effective variables and thus eliminate the problem of multicollinearity. Finally, to locally examine the impact of various factors on catering distribution, the MGWR model (detailed in Section 3.4) was implemented, as well as the OLS regression model and the GWR model for comparison.

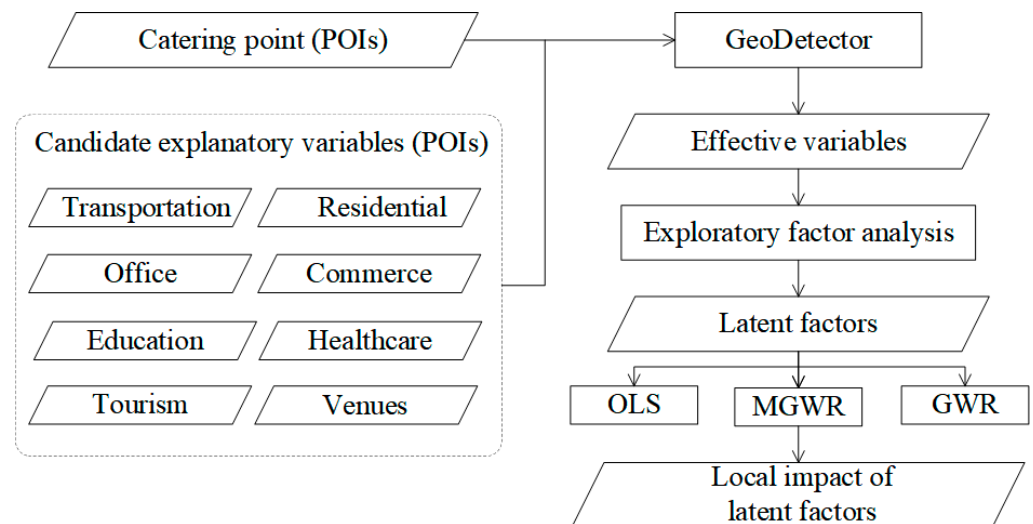


Figure 3. Research framework.

3.2. GeoDetector Analysis

GeoDetector is a spatial statistical method used to quantitatively assess the stratified heterogeneity and measure the spatial correlation between variables [26]. It reveals how variables interact with each other by analyzing their spatial disparities. GeoDetector includes 4 aspects of detection, namely the factor detector, ecological detector, risk detector, and interaction detector. Based on the roles of different detectors, the factor detector and the interaction detector, respectively, were used to detect the intensity and interaction of the variables that contribute to catering distribution.

The factor detector was used to detect the effective variables related to the spatial heterogeneity of catering distribution. Conceptually, spatial stratification or classification is a partition of a target area, where observations are homogeneous within each stratum (zone) but not between strata (zones). The key assumption of GeoDetector is that if variable x is associated with property y , then x and y will exhibit similar spatial distribution and, in other words, the strata of x should tend to be consistent with the strata of y . In view of this, the degree of interpretation of a certain candidate variable x for the catering distribution y can be measured using the q -value, and the calculation formula is as follows:

$$q = 1 - \frac{\sum_{h=1}^L N_h \sigma_h^2}{N \sigma^2}, \quad (2)$$

where L is the number of strata or zones of variable x ; N_h and N are the numbers of units in stratum (zone) h and the entire study area, respectively; σ_h^2 and σ^2 are the variances of property y within stratum (zone) h and the entire study area, respectively. The larger the q -value, the stronger the explanatory power of variable x to property y , where q ($q \in [0, 1]$) indicates that variable x explains ($q \times 100$)% of the spatial distribution of property y .

The interaction detector identifies interactions between different variables, i.e., whether the explanatory powers of different variables x are independent of each other. Furthermore,

the interaction detector is used to evaluate whether the combination of variables x_1 and x_2 increases or decreases the explanatory power for the property y . The detailed principles can be found in the development team's published paper [26,27].

3.3. Exploratory Factor Analysis

As the complex interaction between dependent variables may cause multicollinearity and affect the performance of a linear regression model, EFA was applied to transform the variables into several interpretable factors. EFA is a statistical method based on the common factor model to examine the underlying structure of intercorrelated variables and identify a number of latent factors [28]. Compared to principal component analysis (PCA), EFA is preferable when combined with a factor rotation technique to enhance the interpretability of the extracted components [30].

The common factor model expresses each observed variable $x_j (j = 1, 2 \dots p)$ as a linear combination of m independent common factors ($F_1, F_2 \dots F_m$) and a unique variable u_j :

$$x_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jm}F_m + u_j, \quad (3)$$

where $a_{jk}F_m$ refers to a specific factor loading related to the j^{th} variable on the m^{th} factor F_m .

In this study, a Kaiser–Meyer–Olkin (KMO) test was first performed to measure the sampling adequacy of the dataset (eligibility for factor analysis). Generally, if the KMO value is less than the threshold value of 0.6, the dataset is assumed to be not suited for EFA [28]. Principal component analysis was applied to estimate factor loadings and determine the number of factors that best fit the data. Thereafter, in order to aid factor interpretation, factor rotation was performed with an orthogonal method of varimax. Those factors with higher loadings after rotation were retained for further analysis.

3.4. Multiscale Geographically Weighted Regression

GWR is an approach to deal with the spatially nonstationary relationships between variables by constructing location-specific regression equations [16]. The structure of the classic GWR model is as follows:

$$y_i = \sum_{j=0}^m \beta_j(u_i, v_i)x_{ij} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (4)$$

where y_i is the dependent variable, x_{ij} is the j^{th} independent variable at the i^{th} position, (u_i, v_i) are the coordinates of the i^{th} position, $\beta_j(u_i, v_i)$ is the regression coefficient at the i^{th} position, ε_i is the error term, and m is the number of independent variables.

Given the existence of spatial nonstationarity, GWR is more reasonable in comparison with a global regression, such as the ordinary least squares (OLS) regression. However, GWR assumes that the scale characteristics of spatial variation are the same for all the involved relationships and thus applies a single kernel function and a constant bandwidth for all the processes in the model. However, in many cases, different processes can operate at varying spatial scales. MGWR is an extension of GWR that takes into account the varying spatial scales by using different bandwidths for each dependent variable. The MGWR model is expressed as follows [17,18]:

$$y_i = \sum_{j=0}^m \beta_{bwj}(u_i, v_i)x_{ij} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (5)$$

where β_{bwj} is the bandwidth used for the calibration of the relationship between the j^{th} independent variable and the dependent variable. The bandwidths in the MGWR model indicate the scale at which individual independent variables influence the dependent variable. Smaller bandwidths indicate more local processes, whereas larger bandwidths

indicate more regional processes, with a bandwidth tending to infinity corresponding to the global model.

4. Results

According to the research framework (Figure 3), the GeoDetector method (<http://www.geodetector.cn>) was used to filter out the effective variables. Thereafter, EFA with a KMO test was performed using SPSS software (v 26.0) to identify the latent factors. Three regression models, including OLS, GWR and MGWR, were implemented in MGWR 2.2 (<https://sgsup.asu.edu/sparc/multiscale-gwr#cardBodyresearch-centers> (accessed on 5 February 2024)). For the GWR and MGWR model calibration, the golden-section algorithm was employed to determine the optimal bandwidths based on the minimization of the corrected Akaike information criterion (AICc). An adaptive bi-square spatial kernel was adopted to address the potential edge effects.

4.1. Detection of the Effective Variables

The eight candidate variables in Table 1 were used as independent variables, while catering point density was adopted as the dependent variable to perform the GeoDetector analysis. Each independent variable was divided into 5 categories using the natural breaks (Jenks) method, and units within the same category were considered to belong to a stratum. The q -values for all the eight variables were calculated according to Equation (2). At the same time, the p -value of each variable was also calculated to determine the statistical significance of the results.

As shown in Table 2, the p -value of each candidate variable was less than the significance level of 0.05, and thus the stratification was significant [26]. This confirmed the fact that spatial association existed between the candidate variables and catering point density, which can also be visually observed in Figure 2. Therefore, all the candidate variables were identified as the effective variables with respect to catering distribution in the study area. According to the q statistics, the strength of explanatory power of the eight variables was in descending order of healthcare, education, residential, commerce, office, transportation, venues, and tourism. Moreover, the high values of the variance inflation factor (VIF) of healthcare and education, which were above the threshold of 4, indicated significant multicollinearity between these two variables and the other variables [18].

Table 2. Factor detector and interaction detector results.

Variable	q -Value	p -Value	VIF
Transportation	0.442	0.000	2.402
Residential	0.602	0.000	1.957
Office	0.456	0.000	2.728
Commerce	0.457	0.000	1.600
Education	0.737	0.000	4.421
Healthcare	0.862	0.000	4.355
Tourism	0.337	0.000	1.438
Venues	0.391	0.000	1.360

The interaction detector results revealed the interactions between the eight variables (Figure 4). Compared to the q -values of the individual variables in Table 2, the q -value for the interaction between any two variables was consistently greater than the maximum of the two single variables ($\text{Max}(q(A), q(B)) < q(A \cap B)$), but less than their additive sum ($q(A \cap B) < q(A) + q(B)$). These findings indicated that the explanatory power of any single variable could be enhanced when interacting with another. Consequently, while all the eight variables had a significant impact on catering distribution, the presence of

strong collinearity and interaction between the variables demonstrated the necessity of considering the underlying structure between the variables.

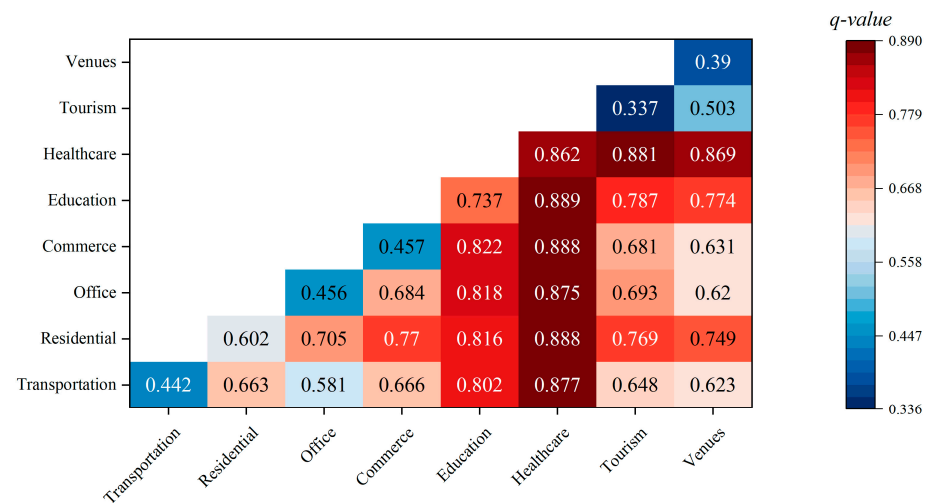


Figure 4. Interaction detection results.

4.2. Extraction of Latent Factors

The high Kaiser–Meyer–Olkin (KMO) value of 0.876 indicated that the data were well-suited for EFA. Table 3 presents the results of factor extraction, and the factor loadings that are higher than 0.4 and lower than −0.4 are highlighted in bold. Four latent factors were retained, with a total variance contribution of 84.746%. Before factor rotation, all the variables initially loaded highly on Factor 1, accounting for 56.048% of the total variance, and several variables were cross-loading. A clearer factor structure was obtained through factor rotation. The variance contribution of each factor became more equally distributed, namely 35.101%, 19.843%, 16.521%, and 13.281%, respectively.

Table 3. Results of factor loadings before and after rotation.

	Variable	Factor 1	Factor 2	Factor 3	Factor 4
Factor loadings before rotation	Transportation	0.767	−0.393	0.136	0.178
	Residential	0.743	−0.317	0.228	−0.070
	Office	0.818	−0.304	0.058	0.128
	Commerce	0.629	0.140	− 0.709	−0.070
	Education	0.897	0.020	−0.050	−0.104
	Healthcare	0.885	0.083	−0.147	−0.166
	Tourism	0.595	0.494	0.392	− 0.417
	Venues	0.580	0.538	0.116	0.592
Variance accounted for before rotation		56.048%	11.355%	9.585%	7.758%
Factor loadings after rotation	Transportation	0.861	0.148	0.044	0.165
	Residential	0.790	0.114	0.268	0.022
	Office	0.818	0.262	0.107	0.182
	Commerce	0.182	0.933	0.049	0.126
	Education	0.618	0.485	0.404	0.196
	Healthcare	0.538	0.583	0.427	0.168
	Tourism	0.179	0.101	0.921	0.185
	Venues	0.192	0.165	0.200	0.941
Variance accounted for after rotation		35.101%	19.843%	16.521%	13.281%

According to the factor loadings after factor rotation, the residential, traffic, and office variables loaded highly onto Factor 1 (F1). The variables of commerce, tourism, and venues loaded extremely highly onto Factor 2 (F2), Factor 3 (F3), and Factor 4 (F4), respectively. Furthermore, the variables of healthcare and education, which showed the strongest interactions with the other variables (see Table 3), loaded equally on F1, F2, and F3.

For each factor, the factor score was calculated as the weighted sum of the variables, using the corresponding factor score coefficient as a weight (see Table 4). The updated VIF values revealed that there was no strong collinearity among the factors present.

Table 4. Factor score coefficient matrix.

Variable	F1	F2	F3	F4
Transportation	0.474	−0.192	−0.258	0.052
Residential	0.400	−0.214	0.082	−0.194
Office	0.386	−0.070	−0.198	0.035
Commerce	−0.245	0.896	−0.204	−0.073
Education	0.095	0.182	0.164	−0.064
Healthcare	0.006	0.316	0.204	−0.118
Tourism	−0.190	−0.166	0.975	−0.106
Venues	−0.118	−0.129	−0.153	1.116
VIF	1.544	1.065	1.528	1.058

4.3. Local Impact of Latent Factors

4.3.1. Model Performance and Comparison

The MGWR model was calibrated using the four factors (F1–F4) extracted through EFA as independent variables, along with catering point density as the dependent variable. For comparison, the performance of the MGWR model was systematically evaluated against both the OLS regression model and the GWR model.

The comparison of the three models' goodness of fit is presented in Table 5. According to the adjusted R^2 , both GWR and MGWR outperformed the OLS model. The adjusted R^2 improved from 0.837 (OLS) to 0.882 (GWR) and 0.903 (MGWR). The MGWR model yielded the highest adjusted R^2 , suggesting that the model could explain more than 90% of the total variance. The residual sum of squares (RSS), indicating unexplained variations, was high in the OLS model (257) but dropped to almost one half in GWR (163) and MGWR (131). Figure 5 displays the maps of residuals from each model, using quantile classification. It was observed that the MGWR residuals were distributed more randomly, showing that more random errors remained. The Moran's I statistic was computed to determine the spatial dependency of the residuals. The OLS and GWR residuals showed significant spatial dependency, with Moran's I values of 0.152 and 0.105 and p -values less than 0.001. The MGWR model, with a Moran's I value of 0.016 and a p -value of 0.355, showed no significant spatial pattern among the residuals. Thus, MGWR not only provided a better fit than OLS and GWR, but also captured the spatial variations in the relationships, removing the dependency in the residuals.

Table 5. Comparison of the OLS, GWR, and MGWR model fits.

Model	R^2	Adjusted R^2	RSS	Moran's I	p -Value	AICc
OLS	0.837	0.837	257.221	0.152	0.000	1627.791
GWR	0.897	0.882	163.108	0.105	0.000	1334.923
MGWR	0.917	0.903	131.200	0.016	0.355	1099.686

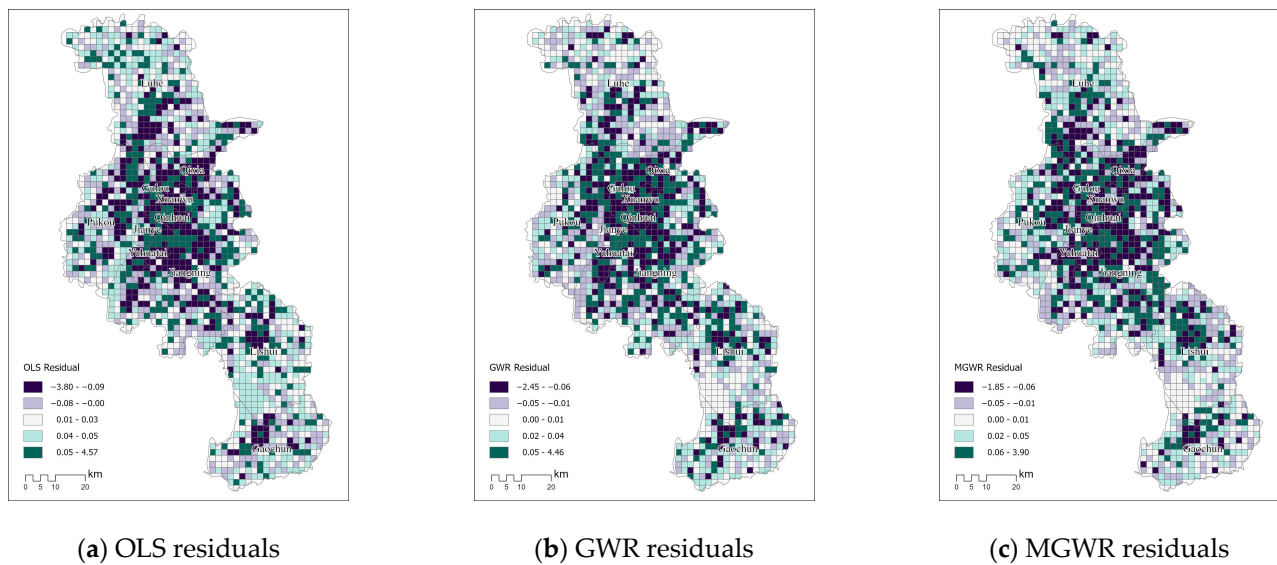


Figure 5. Distribution of local residuals from the OLS model (a), the GWR model (b), and the MGWR model (c).

The AICc represents the relative amount of lost information in the model by considering the risks of both overfitting and underfitting [31]. The OLS showed an AICc of 1626, while the GWR AICc was 1334, and the MGWR further brought the AICc down to 1099. The lowest AICc value of MGWR indicated that more important information was retained, and thus it was a more plausible model.

The performance of the models depended largely on the optimal use of the bandwidth. As shown in Table 6, while the global OLS model considered all the sampling units equally for parameter estimation, the GWR model used a local but single bandwidth of 81, thus considering 81 nearest neighbors ($2 \text{ km} \times 2 \text{ km}$ units) to inform the construction of parameter estimates. In the case of MGWR, instead of assigning a single bandwidth for all the independent variables, it computed the optimal bandwidth for each variable and inferred different spatial scales of relationships between the various variables and catering distribution. The large bandwidth for F3 with 329 nearest neighbors, compared to 30 for F1 and F2 and 97 for F4, indicated that the effect of F3 operated at a broader spatial scale than that of the other factors.

Table 6. Summary statistics and bandwidth of the parameter estimates from the OLS, GWR, and MGWR.

Model	Variable	Min	Median	Max	Bandwidth
OLS	Factor 1		1.086		
	Factor 2		0.424		
	Factor 3		0.432		
	Factor 4		0.068		
	Intercept		0.000		
GWR	Factor 1	−0.251	1.238	2.895	81
	Factor 2	−1.053	0.625	1.595	
	Factor 3	−1.279	0.474	1.964	
	Factor 4	−6.601	0.061	1.540	
	Intercept	−0.935	0.022	0.428	

Table 6. *Cont.*

Model	Variable	Min	Median	Max	Bandwidth
MGWR	Factor 1	0.627	1.053	1.537	30
	Factor 2	−0.182	0.543	1.512	30
	Factor 3	0.275	0.402	0.476	329
	Factor 4	−0.628	0.029	0.622	97
	Intercept	−0.027	−0.025	−0.022	1579

Based on the above information, it can be concluded that the MGWR indeed improved the model performance and allowed for an analysis of the spatially nonstationary effects of various factors on catering distribution.

4.3.2. Local Parameter Estimates

The relationship between the catering point density and the four retained factors (F1–F4) was modeled using the MGWR model. Figure 6 illustrates the spatial variation of the local parameter estimates for each factor at the confidence level of 0.95. The parameter estimates quantified the change in catering point density resulting from a one-unit increase in each factor, with positive values (shown in red) indicating an increase and negative values (shown in blue) indicating a decrease. A larger absolute value of the parameter estimate signified a stronger intensity of the factor's impact. Since all the factors were standardized, the parameter estimates could be directly compared with each other. As shown in the maps, F1, F2, and F3 consistently exhibited positive effects on catering point density (Figure 6a–c), while F4 demonstrated both positive and negative effects across the city (Figure 6d). The varying coloring found in each of the parameter maps indicates that the effects of these factors were not stable over space. Notably, F3 displayed a larger spatial clustering of the parameter estimates compared to the other factors, which aligned with the optimal bandwidth assigned to these factors (refer back to Table 6).

As illustrated in Figure 6a, the local parameter estimates for F1 demonstrated a statistically significant positive relationship with catering point density across the city, with a one-unit increase in F1 corresponding to a maximum 1.5-unit increase in catering point density. Spatial clusters with higher positive cumulative estimates were found in each district. The highest estimate emerged in the central urban area, followed sequentially by Lishui, Luhe, Gaochun, Pukou, and Jiangning. The local estimates for F2, as shown in Figure 6b, were statistically insignificant in the peripheral regions, while they exhibited a positive effect in other regions. This effect was prominent in suburban districts such as Lishui and Pukou, but it diminished in the central urban area. In Figure 6c, the local parameter estimates for F3 revealed a positive relationship across the city. The stronger effects were concentrated in the southern and northern regions, gradually weakening towards the central area. Figure 6d presents a more complex spatial relationship for F4, showing both positive and negative effects across different subregions. While most areas showed no significant relationship, contrasting effects emerged in specific locations. In central Jiangning, F4 acted as a negative estimator, with a one-unit increase in F4 corresponding to a maximum 0.3-unit decrease in catering point density, whereas in northern Jiangning and the central urban area, F4 acted as a positive estimator, with a one-unit increase in F4 corresponding to a maximum 0.5-unit increase in catering point density.

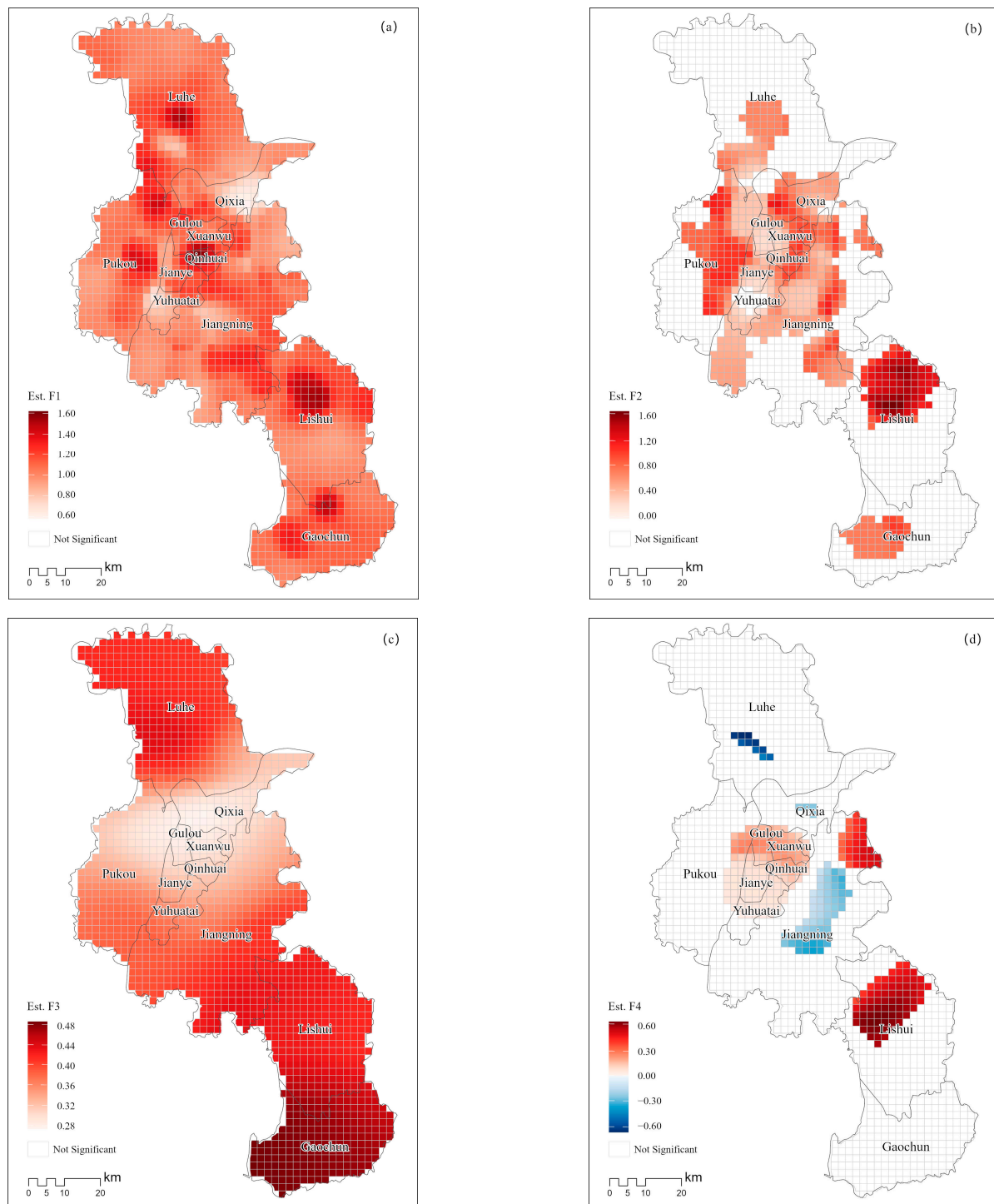


Figure 6. MGWR local parameter estimates for (a) Factor 1, (b) Factor 2 (c), Factor 3, and (d) Factor 4.

5. Discussion

This study used POI data to investigate the factors influencing the distribution of catering points. Eight categories of POIs, namely residential, transportation, office, commerce, education, healthcare, tourism, and venues, were identified as the effective variables affecting catering distribution. The findings underscore the importance of population density, traffic accessibility, commercial areas, infrastructural development, and tourism activities. Similar findings were revealed in previous studies that employed multi-source data such as land use maps, government surveys, and statistical records [5,6,8,9]. However, the inconsistent spatial resolution of data from multiple sources can introduce unexpected

errors. To address this limitation, this study relied exclusively on POI data. Densities of different types of POIs were calculated within uniform statistical units to ensure consistent spatial resolution. Moreover, POI data feature precise location accuracy, real-time capability, and high sensitivity in capturing small changes, making them particularly suitable for exploring the urban environment at fine spatiotemporal resolutions [21,32]. Specifically, we adopted up-to-date traffic POIs (i.e., bus stops, subway stations, and parking lots) rather than traditional road maps to represent traffic accessibility due to their ability to capture both public transportation and automobile accessibility [33]. Notably, empirical studies have proven that parking facilities are critical locational factors for catering points [34]. In addition, compared to census data for measuring population density, residential POIs provide a more accurate and up-to-date population distribution by marking the locations of individual residential building [35,36].

In this study, a novel method incorporating GeoDetector analysis and EFA was proposed to extract effective factors from a set of candidate variables. As emphasized in many previous studies employing regression models, the identification and selection of effective variables are an essential step [37]. The Moran's I statistic revealed spatial heterogeneity in the distribution of both candidate variables and catering points, with Moran's I values ranging from 0.2 to 0.7. Consequently, GeoDetector analysis, which allows for the measurement of spatial dependence, proved to be more suitable for identifying the effective variables than traditional statistical methods such as the correlation coefficient [26]. Furthermore, the results of GeoDetector interaction detection highlight the presence of multicollinearity among the candidate variables. Simply excluding collinear variables could lead to information loss and overlooking the effects of the removed variables [38,39]. EFA, a widely utilized statistical method in the social sciences, was used to address the issue of multicollinearity. By analyzing the shared variance among the variables, EFA effectively explained the correlations between the variables and identified several latent factors [30]. It is worth noting that with the emergence of the big data technology, POI data have greatly enriched the spatial information for urban research with finely sorted categories, but they also increased the potential for data redundancy. The combination of GeoDetector analysis and EFA offers a practical solution to identify effective variables while eliminating redundant data, thereby enhancing the robustness and interpretability of the analysis.

By using EFA, the eight effective variables were transformed into four interpretable latent factors labeled as F1, F2, F3, and F4. Given that the eight variables corresponded to different types of POIs where people conduct daily activities, F1 could be interpreted as catering consumption linked to commuting activities, F2 as shopping activities, F3 as tourism activities, and F4 as gathering activities according to the factor loadings of each variable (refer back to Table 3). It is important to note that healthcare POIs and education POIs, despite exhibiting the strongest effect on catering point distribution, were evenly distributed across all the factors due to significant multicollinearity with the other variables. Consequently, although healthcare and education were not identified as distinct factors, their impact was successfully preserved. It is also worth noting that F1, which had a prominent impact on catering point distribution, indicated a combined effect of the residential, office, and transportation variables. This contrasts with most previous studies, which treated traffic accessibility as a separate factor. In reality, POIs inherently exhibit agglomeration effects, and certain combinations of POIs frequently co-occur [32]. In our case study, the agglomeration of residential, office, and transportation POIs enhanced the emergence of catering points. This finding provides valuable insights for optimizing urban functional spaces in terms of POI configuration patterns.

The spatially nonstationary impacts of various factors were analyzed using the MGWR model, which yielded a better fit than the OLS model and the GWR model. According to the MGWR parameter estimates, all the four factors exerted statistically significant impacts on catering distribution, but the impact magnitude varied considerably across locations. In the central urban area, where the highest density of catering points was observed, F1 (commuting activities) exhibited a strong positive relationship with catering distribution. In contrast, the effects of F2 (shopping activities), F3 (tourism activities), and F4 (gathering activities) were positive but relatively minor. This suggests that within the central urban area, the density of catering points is primarily driven by regular consumption associated with daily commuting activities. Consequently, increasing commercial facilities, tourism attractiveness, or public infrastructure in this area is unlikely to significantly boost the number of catering points. This contrasts with the estimates for the peripheral area of the city. While F1 continued to exert strong positive effects, the impacts of F2 and F3 were remarkably enhanced. This indicates that shopping and tourism activities play a more substantial role in driving the demand for catering points in these areas. Therefore, urban planning efforts aimed at developing new commercial facilities or enhancing tourism attractiveness in peripheral areas should strategically incorporate catering development to meet this demand. Moreover, various bandwidths of the four factors indicate that tourism activities exert regional-scale effects, while commuting and shopping activities exert local-scale effects and gathering activities exert intermediate-scale effects. In conclusion, while previous studies examined the various factors influencing catering point distribution, the results of the MGWR model in this study highlighted that the intensity of impacts varies over space, and the scale of impact differs by individual factors. These findings align with prior research on the relationship between population density and catering distribution [11,32], but provide a more granular intra-urban perspective.

However, this study still has some limitations as follows. First, the scale and the method used to define the statistical unit can significantly affect the results [40]. In this study, grid-based statistical units were constructed following the principles outlined in [29]. Different grid sizes ranging from 0.5 km to 5 km at 0.5 km intervals were attempted, and a 2 km × 2 km grid was identified as the most suitable for the study area and input dataset. Smaller grids may overestimate the relationship between catering distribution and the variables, while larger grids might underestimate the local impact. Given that the optimal grid size is almost at the community level, future studies could implement the model by adopting communities as statistical units, thereby supporting POI configuration among communities. Secondly, POI data, similarly to other big data sources, suffer from possible sampling bias and errors in the crowd-sourced collection process. Notably, informal food services were not included in the catering point dataset, potentially leading to an underestimation of the factor impact. Additionally, the spatial distribution of POIs was calculated based solely on the coordinate information, without considering attributes such as the scale and type of catering points, building height, or the number of transport lines at transportation hubs. More accurate results could be obtained by incorporating these facility attributes. Finally, given the dynamic nature of POI data, future research should investigate both spatial and temporal nonstationarity in the relationships between various factors and catering distribution. This would provide more comprehensive insights into catering location and the optimization of urban functional spaces.

6. Conclusions

This study investigated the spatially nonstationary impact of various factors on catering point distribution. Using the POI data, interpretable latent factors were extracted through a novel method incorporating GeoDetector analysis and EFA and examined locally

by implementing the MGWR model. Taking the city of Nanjing, China, as a case study, four factors having a significant effect on catering distribution were identified using POI data from Baidu Maps, which could be interpreted as commuting activities, shopping activities, tourism activities, and gathering activities according to the daily activities conducted at the corresponding POIs. The MGWR model, in comparison with the OLS regression model and the GWR model, was applied to explore the relationship between the various factors and catering point distribution. Our results confirm that the MGWR model, by adopting variable bandwidths, can achieve a higher goodness of fit in the context of spatial variation. The parameter estimates of the MGWR model reveal the extent to which each factor would affect catering point distribution at a specific location. The findings provide a valuable reference for guiding catering development and optimizing urban functional spaces.

Author Contributions: Conceptualization, Lu Tan and Xiaojun Bu; methodology, Lu Tan and Xiaojun Bu; software, Xiaojun Bu; validation, Lu Tan; investigation, Lu Tan; resources, Xiaojun Bu; writing—original draft preparation, Xiaojun Bu; writing—review and editing, Lu Tan; visualization, Lu Tan and Xiaojun Bu; funding acquisition, Lu Tan. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Natural Science Foundation of Nanjing Forestry University grant number CX2016005.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s).

Acknowledgments: We would like to thank the anonymous reviewers for taking the time and effort to review the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fisher, D.P. Location, location, location: Ensuring a franchisee's success. *Hosp. Rev.* **1997**, *15*, 4.
2. Jung, S.S.; Jang, S.S. To cluster or not to cluster?: Understanding geographic clustering by restaurant segment. *Int. J. Hosp. Manag.* **2019**, *77*, 448–457. [[CrossRef](#)]
3. Qin, X.; Zhen, F.; Gong, Y. Combination of big and small data: Empirical study on the distribution and factors of catering space popularity in Nanjing, China. *J. Urban Plan. Dev.* **2019**, *145*, 05018022. [[CrossRef](#)]
4. Mulligan, G.F. Agglomeration and central place theory: A review of the literature. *Int. Reg. Sci. Rev.* **1984**, *9*, 1–42. [[CrossRef](#)]
5. Austin, S.B.; Melly, S.J.; Sanchez, B.N.; Patel, A.; Buka, S.; Gortmaker, S.L. Clustering of fast-food restaurants around schools: A novel application of spatial statistics to the study of food environments. *Am. J. Public Health* **2005**, *95*, 1575–1581. [[CrossRef](#)]
6. Liang, L. The distribution in space of urban catering and its factors: Xi'an as an example. *J. Northwest Univ. (Nat. Sci. Ed.)* **2007**, *37*, 925–930.
7. Litz, R.A.; Rajaguru, G. Does small store location matter? A test of three classic theories of retail location. *J. Small Bus. Entrep.* **2008**, *21*, 477–492. [[CrossRef](#)]
8. Melaniphy, J.C. *Restaurant and Fast Food Site Selection*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1992.
9. Shu, S.Y.; Wang, R.; Sun, W.Y.; Liu, J.; Xiao, L.S. Spatial distribution of urban catering industry and its influenced factors: A case study of Xiamen City. *Trop. Geogr.* **2012**, *32*, 134–140.
10. Zhang, X.; Xu, Y. Study on the distribution in space of urban caterings and its influencing factors: A case study of Nanjing. *Trop. Geogr.* **2009**, *29*, 134–140.
11. Ma, D.; Liu, B.; Huang, Q.; Zhang, Q. Evolution characteristics and causes—An analysis of urban catering cluster spatial structure. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 302. [[CrossRef](#)]
12. Chen, Y.; Yin, G.; Hou, Y. Street centrality and vitality of a healthy catering industry: A case study of Jinan, China. *Front. Public Health* **2022**, *10*, 1032668. [[CrossRef](#)] [[PubMed](#)]
13. Wu, M.; Pei, T.; Wang, W.L.; Guo, S.H.; Song, C.; Chen, J.; Zhou, C.H. Roles of locational factors in the rise and fall of restaurants: A case study of Beijing with POI data. *Cities* **2021**, *113*, 103185. [[CrossRef](#)]
14. Shabrina, Z.; Buyuklieva, B.; Ng, M.K.M. Short-term rental platform in the urban tourism context: A geographically weighted regression (GWR) and a multiscale GWR (MGWR) approaches. *Geogr. Anal.* **2021**, *53*, 686–707. [[CrossRef](#)]

15. Wang, W.W.; Wang, S.; Chen, H.; Liu, L.J.; Fu, T.L.; Yang, Y.X. Analysis of the characteristics and spatial pattern of the catering industry in the four central cities of the Yangtze River Delta. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 321. [\[CrossRef\]](#)
16. Brunsdon, C.; Fotheringham, A.S.; Charlton, M.E. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geogr. Anal.* **1996**, *28*, 281–298. [\[CrossRef\]](#)
17. Fotheringham, A.S.; Yang, W.; Kang, W. Multiscale geographically weighted regression (MGWR). *Ann. Am. Assoc. Geogr.* **2017**, *107*, 1247–1265. [\[CrossRef\]](#)
18. Yu, H.C.; Fotheringham, A.S.; Li, Z.Q.; Oshan, T.; Kang, W.; Wolf, L.J. Inference in multiscale geographically weighted regression. *Geogr. Anal.* **2020**, *52*, 87–106. [\[CrossRef\]](#)
19. Dong, L.; Ratti, C.; Zheng, S. Predicting neighborhoods' socioeconomic attributes using restaurant data. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 15447–15452. [\[CrossRef\]](#)
20. Einav, L.; Levin, J. Economics in the age of big data. *Science* **2014**, *346*, 1243089. [\[CrossRef\]](#)
21. Psyllidis, A.; Gao, S.; Hu, Y.; Kim, E.K.; McKenzie, G.; Purves, R.; Yuan, M.; Andris, C. Points of Interest (POI): A commentary on the state of the art, challenges, and prospects for the future. *Comput. Urban Sci.* **2022**, *2*, 20. [\[CrossRef\]](#)
22. Wu, C.; Ye, X.Y.; Ren, F.; Du, Q.Y. Check-in behaviour and spatio-temporal vibrancy: An exploratory analysis in Shenzhen, China. *Cities* **2018**, *77*, 104–116. [\[CrossRef\]](#)
23. Li, Y.; Liu, H.; Wang, L.E. Spatial distribution pattern of the catering industry in a tourist city: Taking Lhasa city as a case. *J. Res. Ecol.* **2020**, *11*, 191–205.
24. Zhang, Y.; Min, J.; Liu, C.X.; Li, Y.C. Hotspot detection and spatiotemporal evolution of catering service grade in mountainous cities from the perspective of geo-information tupu. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 287. [\[CrossRef\]](#)
25. Zhou, Y.; Shen, X.; Wang, C.; Liao, Y.X.; Li, J.L. Mining the spatial distribution pattern of the typical fast-food industry based on point-of-interest data: The case study of Hangzhou, China. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 559. [\[CrossRef\]](#)
26. Wang, J.F.; Zhang, T.L.; Fu, B.J. A measure of spatial stratified heterogeneity. *Ecol. Indic.* **2016**, *67*, 250–256. [\[CrossRef\]](#)
27. Wang, J.F.; Li, X.H.; Christakos, G.; Liao, Y.L.; Zhang, T.; Gu, X.; Zheng, X.Y. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 107–127. [\[CrossRef\]](#)
28. Howard, M.C. A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *Int. J. Hum. Comput. Interact.* **2016**, *32*, 51–62. [\[CrossRef\]](#)
29. Hengl, T. Finding the right pixel size. *Comput. Geosci.* **2006**, *32*, 1283–1298. [\[CrossRef\]](#)
30. Costello, A.B.; Osborne, J. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Pract. Assess. Res. Evaluation* **2019**, *10*, 7.
31. Akaike, H. Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*; Parzen, E., Tanabe, K., Kitagawa, G., Eds.; Springer: New York, NY, USA, 1998; pp. 199–213.
32. Liu, K.; Yin, L.; Lu, F.; Mou, N.X. Visualizing and exploring POI configurations of urban regions on POI-type semantic space. *Cities* **2020**, *99*, 102610. [\[CrossRef\]](#)
33. Zhou, J.; Yang, Y. Transit-based accessibility and urban development: An exploratory study of Shenzhen based on big and/or open data. *Cities* **2021**, *110*, 102990. [\[CrossRef\]](#)
34. Tzeng, G.H.; Teng, M.H.; Chen, J.J.; Opricovic, S. Multicriteria selection for a restaurant location in Taipei. *Int. J. Hosp. Manag.* **2002**, *21*, 171–187. [\[CrossRef\]](#)
35. Li, J.G.; Li, J.W.; Yuan, Y.Z.; Li, G.F. Spatiotemporal distribution characteristics and mechanism analysis of urban population density: A case of Xi'an, Shanxi, China. *Cities* **2019**, *86*, 62–70. [\[CrossRef\]](#)
36. Zhao, Y.; Li, Q.; Zhang, Y.; Du, X. Improving the accuracy of fine-grained population mapping using population-sensitive POIs. *Remote Sens.* **2019**, *11*, 2502. [\[CrossRef\]](#)
37. Zhao, R.; Zhan, L.P.; Yao, M.X.; Yang, L.C. A geographically weighted regression model augmented by Geodetector analysis and principal component analysis for the spatial distribution of PM_{2.5}. *Sustain. Cities Soc.* **2020**, *56*, 102106. [\[CrossRef\]](#)
38. Nguyen, V.C.; Ng, C.T. Variable selection under multicollinearity using modified log penalty. *J. Appl. Statl.* **2019**, *47*, 201–230. [\[CrossRef\]](#)
39. Tsao, M. Estimable group effects for strongly correlated variables in linear models. *Stat. Plan. Infer.* **2019**, *198*, 29–42. [\[CrossRef\]](#)
40. Dark, S.J.; Bram, D. The modifiable areal unit problem (MAUP) in physical geography. *Prog. Phys. Geog.* **2007**, *31*, 471–479. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.