



Article

Spatial Distribution Characteristics and Driving Factors of Little Giant Enterprises in China's Megacity Clusters Based on Random Forest and MGWR

Jianshu Duan , Zhengxu Zhao, Youheng Xu, Xiangting You, Feifan Yang and Gang Chen * 

Department of Geographic Information Science, School of Geographic and Oceanographic Sciences, Nanjing University, Nanjing 210023, China; 502022270063@smail.nju.edu.cn (J.D.); 502023270114@smail.nju.edu.cn (Z.Z.); 502023270100@smail.nju.edu.cn (Y.X.); xiangting_you@smail.nju.edu.cn (X.Y.); feifanyang@smail.nju.edu.cn (F.Y.)
* Correspondence: chengang@nju.edu.cn

Abstract: As a representative of potential “hidden champions”, a concept originating in Germany, specialized and innovative Little Giant Enterprises (LGEs) have become exemplary models for small and medium-sized enterprises (SMEs) in China. These enterprises are regarded as crucial support for realizing the strategy of building a strong manufacturing country and addressing the weaknesses in key industrial areas. This paper begins by examining urban agglomerations, which serve as the main spatial carriers for industrial restructuring and high-quality development in manufacturing. Based on data from LGEs in the Yangtze River Delta (YRD) and Pearl River Delta (PRD) urban agglomerations from 2019 to 2023, the study employs the Random Forest (RF) and Multi-scale Geographically Weighted Regression (MGWR) methods to conduct a comparative analysis of their spatial patterns and influencing factors. The results are as follows: (1) LGEs exhibit spatial clustering in both the YRD and PRD regions. Enterprises in the YRD form a “one-axis-three-core” pattern within a distance of 65 km, while enterprises in the PRD present a “single-axis” pattern within a distance of 30 km, with overall high clustering intensity. (2) The YRD is dominated by traditional manufacturing and supplemented by high-tech services. In contrast, the PRD has a balanced development of high-tech manufacturing and services. Enterprises in different industries are generally characterized by a “multi-point clustering” characteristic, of which the YRD displays a multi-patch distribution and the PRD a point–pole distribution. (3) Factors such as industrial structure, industrial platforms, and logistics levels significantly affect enterprise clustering and exhibit scale effects differences between the two urban clusters. Factors such as industrial platforms, logistics levels, and dependence on foreign trade show positive impacts, while government fiscal expenditure shows a negative impact. Natural geographical location factors exhibit opposite effects in the two regions but are not the primary determinants of enterprise distribution. Each region should leverage its own strengths, improve urban coordination and communication mechanisms within the urban cluster, strengthen the coordination and linkage of the manufacturing industry chain upstream and downstream, and promote high-tech industries, thereby enhancing economic resilience and regional competitiveness.

Keywords: specialized and innovative little giant enterprises (LGEs); urban agglomeration; random forest; multiple geographically weighted regression; YRD; PRD



Citation: Duan, J.; Zhao, Z.; Xu, Y.; You, X.; Yang, F.; Chen, G. Spatial Distribution Characteristics and Driving Factors of Little Giant Enterprises in China's Megacity Clusters Based on Random Forest and MGWR. *Land* **2024**, *13*, 1105. <https://doi.org/10.3390/land13071105>

Academic Editors: Wei Lang, Tingting Chen, Eddie C.M. Hui and Xun Li

Received: 25 June 2024

Revised: 17 July 2024

Accepted: 19 July 2024

Published: 22 July 2024



Copyright: © 2024 by the authors. 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

Hermann Simon, in analyzing the successful experiences of the German export trade, found that local small and medium-sized enterprises (SMEs), especially those leading in international markets, play a significant role in Germany's export trade [1]. From this observation, he introduced the concept of “hidden champions”. This is similar to the specialized and innovative Little Giant Enterprises (referred to hereinafter as LGEs) proposed in China's 2011 issuance of the “Twelfth Five-Year Plan for the Growth of Small

and Medium-sized Enterprises” [2]. Specialized and innovative enterprises are SMEs that simultaneously possess specialization [3], refinement [4], distinctive characteristics, and outstanding innovation capabilities [3,5]. The most outstanding of these enterprises are referred to as LGEs [6]. In recent years, China has cultivated a cumulative total of 124,000 specialized and innovative SMEs, including 12,000 LGEs [7]. Among these LGEs, 90% serve as ancillary suppliers to well-known large domestic and international corporations, with their research and development intensity being 1.66 times the market average [8]. Urban agglomerations, as carriers for establishing collaborative innovation frameworks among enterprises, provide a significant reference value for promoting the transformation, upgrading, and innovation-driven development of Chinese SMEs by studying the spatial distribution mechanism differences of LGEs within these urban clusters [9].

LGEs represent typical innovative SMEs. Despite being influenced by factors such as national strategic direction and market selection uncertainties, they exhibit high levels of spatial clustering and localization characteristics [10]. Industrial agglomerations have always been an academic hot spot, with many scholars proposing well-known clustering theories such as the Diamond Model [11], Core–Periphery Model [12], Learning Region Theory [13], Tacit Knowledge Approach [14], and Network Approach [15]. However, the above studies are constrained to the continuous agglomeration and self-reinforcement of a single industry in a single region [16]. As research progresses deeper, a more prevalent phenomenon in practical economic activities is inter-industry synergistic clustering. Further exploration reveals the driving factors behind this, as emphasized by Marshall’s external economies theory, including intermediate product sharing, labor pooling, and knowledge diffusion [17]. With the deepening trends of globalization and digitalization, the development and clustering of LGEs are increasingly influenced not only by local factors but also significantly by global market dynamics and changes in international supply chains. Therefore, studying such enterprises requires consideration of more complex economic environments and multi-layered geographical factors.

Currently, there have been numerous studies focusing on such enterprises, primarily concentrated in two areas: first, research on enterprise development and related policies, including policy evolution [18], growth paths [19], and transformation directions [20]; and secondly, analyses similar to this paper’s theme, examining the spatial distribution characteristics and influencing factors of enterprises. Scholars have conducted studies on the spatial distribution patterns of enterprises based on macro administrative divisions, such as urban agglomerations or provincial levels [21–23], and the results indicate spatial heterogeneity in the distribution patterns of enterprises among cities. Specifically, LGEs are mainly concentrated in the eastern coastal, southern coastal, and middle reaches of the Yangtze River regions of China, significantly influenced by the local level of economic development [24]. In recent years, with the accessibility of enterprise information and the application of big data [25], spatial analyses of enterprises have often been conducted at the city scale [26], exploring the complex causal mechanisms driving the emergence of high-growth entrepreneurial enterprises in urban business environments. In the process of enterprise growth and cultivation, favorable natural geographical conditions can provide abundant natural resources and suitable climatic conditions for enterprise development and industrial agglomeration [27]. Transportation infrastructure is also crucial for enterprise development, with regional economics exploring early on how transportation costs affect enterprise location choices. Better transportation accessibility facilitates lower transportation costs and attracts enterprise agglomeration [28]. As SMEs, LGEs find it challenging to bear high land-use costs [29,30], so the availability, price, and related policy support of land influence enterprise location decisions. Scientific research and innovation conditions are core to the development of LGEs [31,32], with their prospects dependent on achieving scale effects, sharing effects, and industry policy support [33]. The resolution of their development challenges relies on external support conditions for the enterprises [34,35]. Additionally, some scholars have considered the perspective of laborers, examining the completeness of surrounding service facilities and their impact on lifestyles [36,37].

From the perspective of research methods, studies on spatial patterns often employ GIS spatial visualization techniques such as spatial autocorrelation and kernel density analysis to analyze their characteristics. For investigating influencing factors, the number of enterprises within grids or administrative divisions is commonly used as the dependent variable. In terms of specific research methods, geographical detectors [38], which can explore the differences and interactions among influencing factors, are frequently utilized. These methods help mitigate the estimation inconsistencies arising from the non-normal distribution of the dependent variable, unlike truncated regression or negative binomial regression [39,40]. With the further development of economic geography, geographically weighted regression and multiscale geographically weighted regression models have been applied in such studies [41,42], focusing more on the heterogeneity of influencing factors at the regional level.

However, the aforementioned studies have certain limitations concerning the focus of this research. Specifically, current studies predominantly concentrate on individual regions, focusing solely on the distribution characteristics and influencing factors within those regions, without considering inter-regional differences. Given China's vast territory and significant spatial heterogeneity, restricting studies to a single city or urban agglomeration results in a lack of fine-grained spatial analysis and fails to capture the impacts of economic, policy, and transportation integration among cities. In selecting research subjects, existing studies have overlooked the industry dimension differences of LGEs, making it difficult to depict the full scope of these enterprises and diminishing the practical significance of the findings. Moreover, previous research has described all significant influencing factors without considering their relative importance, thus failing to highlight the key aspects of the transformation process of SMEs into LGEs.

Building LGEs is a crucial driver for promoting the high-quality development of SMEs in China. It is also an inevitable process to realize the new development pattern of "with domestic circulation as the mainstay, domestic-international dual circulation reinforcing each other", which focuses on the domestic economy while promoting mutual reinforcement between domestic and international economies. Furthermore, focusing on the sustainable development paths of LGEs not only enhances regional economic resilience [43] but also promotes ecosystem health and regional ecological security [44], ensuring a balance between economic growth and environmental protection [45]. Therefore, studying the spatial distribution and influencing factors of LGEs provides valuable insights for promoting the transformation, upgrading, and innovation-driven development of SMEs in China. This paper utilizes data on LGEs from China's Ministry of Industry and Information Technology, focusing on the YRD and the PRD regions. By comprehensively employing geographic big data, county-level yearbook data, and GIS spatial analysis methods, the research investigates the spatial distribution pattern differences of LGEs between these two regions. After employing Random Forest (RF) to filter variable importance, Multi-scale Geographically Weighted Regression (MGWR) is used to explore the correlation and spatial heterogeneity of economic, policy, and ecological driving factors on the distribution of LGEs. The findings aim to provide strategic recommendations for industrial transformation and upgrading, the cultivation of LGEs, and sustainable economic development.

2. Materials and Methods

2.1. Study Area and Data

2.1.1. Study Area Overview

The YRD region is located in East China and serves as a crucial intersection for the Belt and Road Initiative and the Yangtze River Economic Belt, holding significant strategic importance. Conversely, the PRD is situated along the southern coast of China and has seen deepening regional cooperation under the development plans of the Belt and Road Initiative and the Guangdong–Hong Kong–Macao Greater Bay Area. Both regions are characterized as the most open and economically vibrant in China, leveraging rich innovation resources, robust technological innovation capabilities, and favorable business

environments to become hubs for specialized and innovative enterprises, including LGEs. Based on regional planning outlines for the YRD and PRD (Table 1) and considering data availability, this study selects 16 core cities in the YRD and 9 cities in the PRD as the study areas (Figure 1). Through data analysis and comparison, it is noted that the 16 core cities in the YRD account for 75.1% of the LGEs across the entire YRD, making them representative core areas.

Table 1. The scope of the study area.

Name of Megalopolis	Province	Cities Included
Yangtze River Delta Cities	Shanghai	Shanghai
	Jiangsu	Nanjing, Suzhou, Wuxi, Changzhou, Zhenjiang, Yangzhou, Taizhou, Nantong
	Zhejiang	Hangzhou, Ningbo, Huzhou, Jiaxing, Shaoxing, Zhoushan
Pearl River Delta Cities	Guangdong	Shenzhen, Guangzhou, Foshan, Zhongshan, Zhuhai, Dongguan, Huizhou, Zhaoqing, Jiangmen

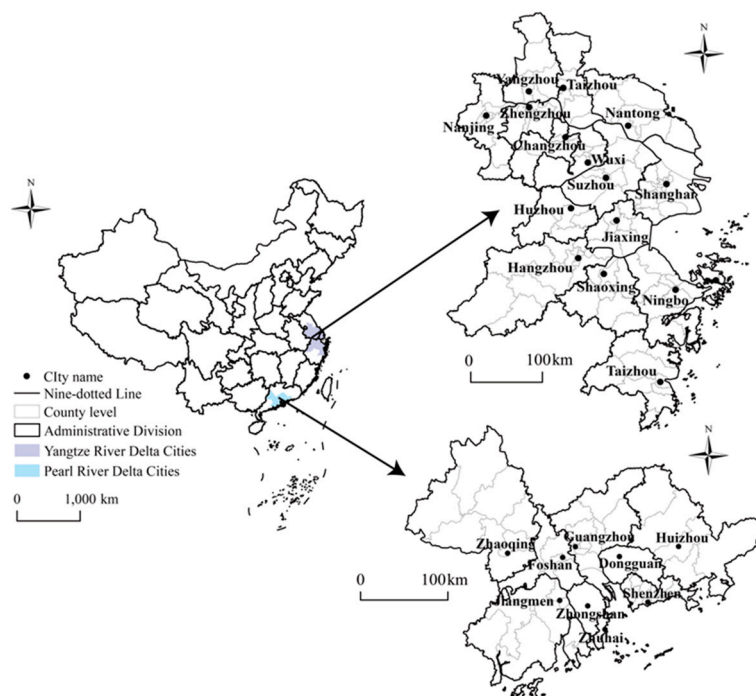


Figure 1. Sketch map of study area.

2.1.2. Research Data

The data utilized in this study primarily consist of two parts: national-level location data of specialized and innovative LGEs situated in the YRD and PRD regions, and data pertaining to the influencing factors. The LGE location data are sourced from the five batches of lists of specialized and innovative LGEs published by provincial (city, district) departments of industry and information technology from 2019 to 2023. These lists were cross-referenced and matched with data from Enterprise Search and the Baidu Maps API, followed by geospatial coordinate system correction and data cleaning processes. This yielded a total of 3241 LGEs in the YRD region and 1488 LGEs in the PRD Area after data processing (Figure 2).

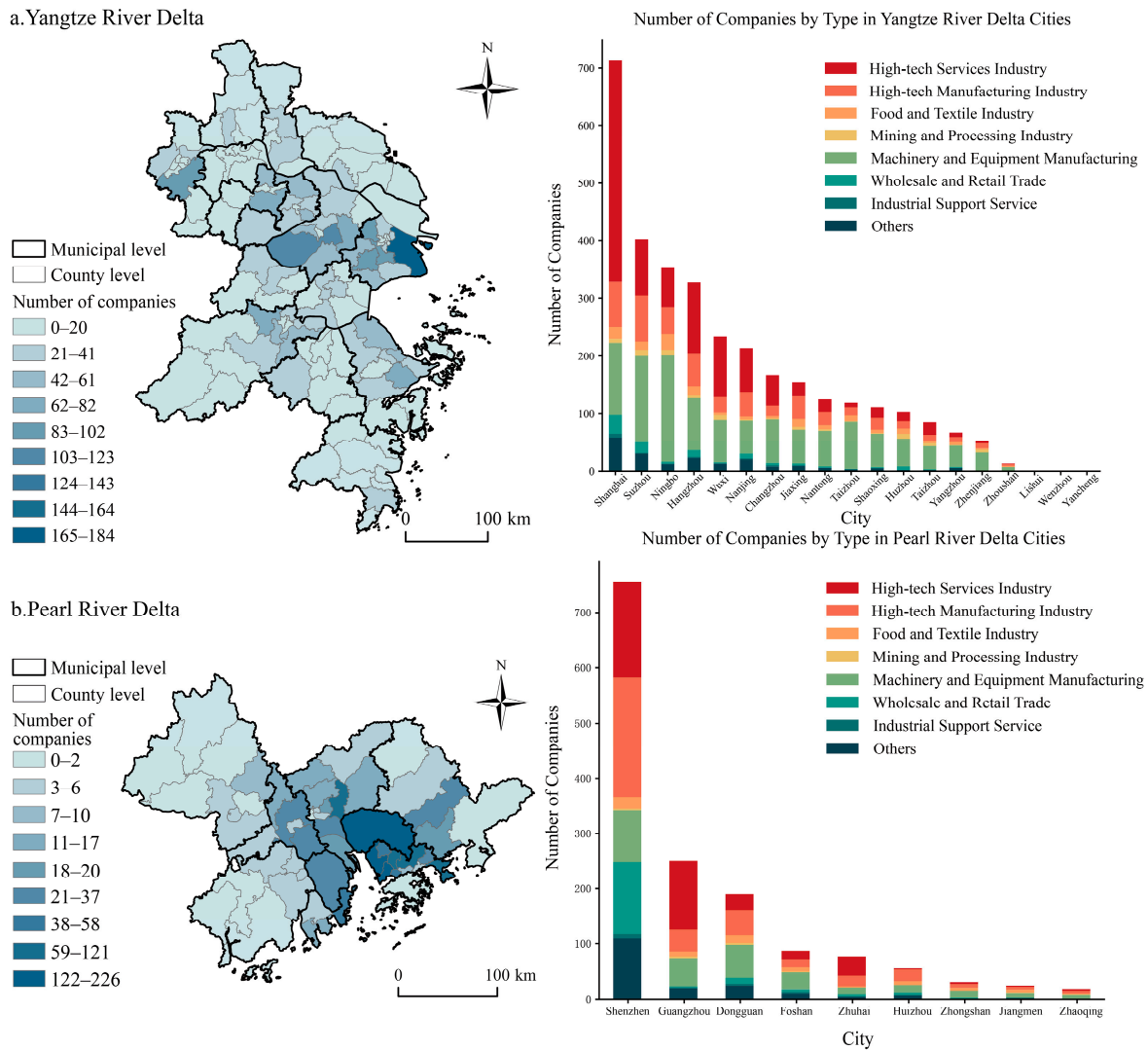


Figure 2. The Distribution of Enterprises in the YRD and PRD.

Based on the current status and characteristics of industrial development, and referring to the National Economic Industry Classification (GB/T 4754-2017) standard, as well as industry classification methods from studies by Ding Jianjun et al. [23], industries are classified into eight major categories (Appendix A). Furthermore, a grid-based approach is employed to process the YRD and PRD regions. To balance sample size and spatial heterogeneity, a grid size of 5 km × 5 km is selected, and a partitioned statistical method is used to calculate the number of LGEs within each grid. The influencing factor data are shown in Table 2 and consist mainly of seven aspects, namely, natural geography and location [27], transportation accessibility [28], land use and cost [29,30], living convenience [36,37], scientific research and innovation conditions [31,32], industrial development basis [33], and external supporting conditions [34,35]. These data are sourced from the National Bureau of Statistics, local government official websites, the National Geospatial Information Center, and other related sources.

Table 2. Selection of influencing factors for distribution of LGEs.

Impact (Level 1) Dimension	Influencing Factors	Variable Description	Data Source
Natural geography and location	Altitude	Average altitude within the grid	NESSDC
	Hydrophilism	Logarithm of the distance to the nearest water body	OSM
	Central area	Distance to Urban Built-up Area	Esri_Land_Cover
Transportation accessibility	Short-distance transport accessibility	Number of bus stops within the grid	AMAP
	Instant transportation accessibility	Number of subway stations in the grid	AMAP
	Medium- and long-distance transportation accessibility	Distance from the center of the grid to the nearest toll station	AMAP
	Road network density	Logarithm of the total length of the road network within the grid	OSM
Land use and cost	Degree of land use	Proportion of urban construction land area in the grid	RESDC
	New home housing costs	Average price of new houses in the grid	Juhui Data Network
	Second-hand housing costs	Average price of second-hand houses in the grid	Juhui Data Network
	Average listing price of the community	Average price of new houses in the grid	Anjuke
Living convenience	Residential convenience	Number of residential communities within the grid	AMAP
	Vehicle carrying capacity	Number of parking lots within the grid	AMAP
Scientific research and innovation conditions	Collaborative innovation basis	Number of higher education institutions within the grid	AMAP
		Number of research institutions within the grid	AMAP
	Human capital level	Number of undergraduate and college students/Total population of the region	Local Statistical Yearbook

Table 2. Cont.

Impact (Level 1) Dimension	Influencing Factors	Variable Description	Data Source
Industrial development basis	Industrial platforms	Number of industrial parks within the grid	AMAP
	Industrial structure	Proportion of the added value of the secondary and tertiary industries to the regional GDP	Local Statistical Yearbook
	Labor market	Average population density within the grid	Landscan Global Population Database
External supporting conditions	Logistics level	Road freight volume/Area of the administrative district	Local Statistical Yearbook
	Accessibility of credit resources	Number of banks in the grid	AMAP
	Government fiscal expenditure ratio	Proportion of general public budget expenditure to regional GDP	Local Statistical Yearbook
	Foreign trade dependency	Exports volume/GDP	Local Statistical Yearbook
	Development zone policies	Number of development zones within the grid	Official websites of the provincial and municipal departments of industry and technology

Note: NESSDC (<http://www.geodata.cn/>) accessed on 25 April 2024, OSM (<http://www.osm.org/>) accessed on 26 April 2024, Esri_Land_Cover (<https://livingatlas.arcgis.com/landcover/>), AMAP (<https://lbs.amap.com/>) accessed on 28 April 2024, RESDC (<https://www.resdc.cn/>) accessed on 20 April 2024, Juhui Data Network (<https://fangjia.gotohui.com/xtop/>), Anjuka (<https://www.anjuka.com/>) accessed on 18 April 2024, Landscan (<https://landscan.ornl.gov/>), accessed on 1 May 2024.

2.2. Research Framework

The overall research framework is shown in Figure 3, which is divided into three main parts: description of data sources and classification of influencing factors, construction of the feature matrix, and analysis of spatial distribution patterns and influencing factors. First, the analysis of the spatial distribution pattern of LGEs is divided into two parts: overall industry analysis and sub-industry analysis. In the overall industry analysis, spatial autocorrelation analysis is used, followed by the application of Ripley's K to determine the optimal aggregation scales of the two regions, which is further used as the bandwidth parameter for kernel density estimation, to conduct kernel density estimation analysis and identify hot and cold spots areas, and to analyze their differences, and then, constructs the feature matrix through the results obtained. The sub-industry analysis compares the spatial distribution of the kernel density of the number of enterprises to reveal the spatial distribution characteristics of enterprises in each industry, thus comparing the variability of the spatial distribution of enterprises in each industry in the two study regions.

Secondly, the LGE influence factors were analyzed as follows: firstly, the influence factor data were preprocessed, the missing values and outliers were processed and eliminated, the variance inflation factor (VIF) was calculated and the variables with VIF greater than 10 were eliminated, and finally, 16 variables in each of the two regions were screened out. Then, the variables were ranked in order of importance using the RF method, and the top ten influencing factors in each of the two regions were screened out. Finally, the

MGWR model is used to explore the positive and negative correlations (coefficient means) of these key influencing factors on the distribution of LGEs and their spatial heterogeneity, revealing the strength and direction of the role of each factor in different regions.

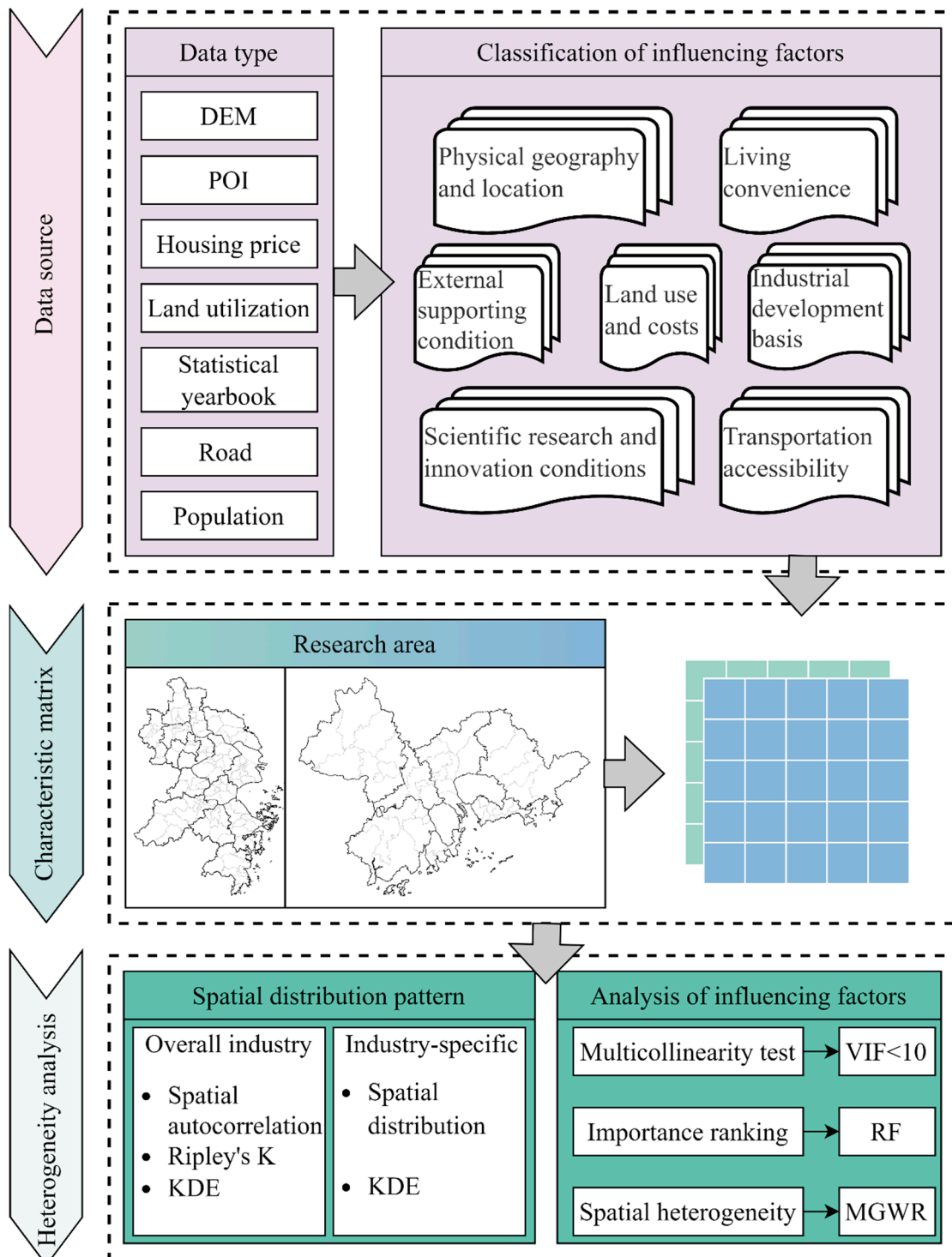


Figure 3. Flowchart of the research framework.

2.3. Methodology

2.3.1. Spatial Autocorrelation

Spatial autocorrelation is widespread in a variety of geographic phenomena [46]. In this study, the global spatial autocorrelation G-statistic and the local spatial autocorrelation

G-statistic are used to analyze the spatial distribution patterns of LGEs in the YRD and PRD regions. The global G-statistic is used to measure whether LGEs in the entire study area are characterized by a pattern of high-value clustering, random, or low-value clustering at the global level [47]. The local G-statistic is used to identify significant hot spots and cold spots of LGEs within specific local areas and to determine the exact location of high- and low-value clusters [48].

2.3.2. Ripley's K Function

Ripley's K function is commonly used to explore how the spatial distribution aggregation of point features changes with distance, and how it varies with changes in neighborhood size. It is valued for its accuracy, simplicity, and ease of use, making it suitable for multi-scale spatial pattern analysis [49]. It is particularly applicable to the point data of LGEs in the YRD and PRD regions. In this study, we employ Ripley's K function to analyze the spatial clustering characteristics of LGEs at different distance scales and identify the peak clustering distances.

2.3.3. Kernel Density Analysis

Kernel density, by calculating the density of an element in its surrounding neighborhood [50], can reflect the degree of agglomeration of the element in space. This study employs kernel density analysis to examine the spatial clustering characteristics of LGEs in the YRD and PRD regions, both overall and within specific industries. The calculation formula is as follows:

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n k \left[\frac{(x - x_i)}{h} \right] \quad (1)$$

where $f(x)$ represents the kernel density estimate, also known as the kernel function, h is the bandwidth parameter, n denotes the number of LGEs within the bandwidth h , and $x - x_i$ denotes the distance from the evaluation point x to the data point x_i [51].

2.3.4. Random Forest Regression Model

This study uses the Random Forest (RF) algorithm to analyze the main drivers of the spatial distribution of LGEs in the YRD and PRD regions and assesses the importance of various factors influencing their distribution [52,53].

RF is a regression method of ensemble learning [54], composed of multiple unpruned regression trees with maximum depth. The final prediction is obtained by averaging the predictions of each tree. In contrast to standard regression, RF does not rely on strict statistical assumptions and is able to model complex correlations and consider interactions between variables. The spatial distribution of the LGEs is used as the dependent variable, and each influencing factor is input into the model as an independent variable, using 100 trees with five variables randomly sampled at each split for training. The importance of each variable was calculated through the Gini coefficient, which reflects the frequency of the variable as a split node and its contribution to model error reduction. Variable Importance Measure (VIMP) is used to identify the primary factors significantly influencing the spatial distribution of LGEs. Model performance is evaluated using the coefficient of determination (R^2) and standard error (SE), and the robustness and reliability of the model are validated through cross-validation methods [55,56].

2.3.5. Multi-Scale Geographically Weighted Regression

Multi-scale geographically weighted regression (MGWR) can be used to study the correlation between spatially distinct explanatory and dependent variables [57]. Traditional geographically weighted regression (GWR) models use a separate unique bandwidth parameter to control the distance to each sampling point when spatially weighting [58] and consider all spatial variables operating at the same scale. In contrast, MGWR relaxes the assumption of a single bandwidth by allowing different variables to have varying bandwidth ranges. This allows relationships between the dependent variable and different

explanatory variables to vary across different spatial scales [3]. Consequently, the MGWR model minimizes overfitting, reduces estimation bias, and mitigates multicollinearity issues in the model, thereby significantly enhancing its predictive performance. MGWR is expressed as follows:

$$Y_i = \sum_{j=1}^k \beta_{bw_j}(u_i, v_i) X_{ij} + \varepsilon_i \quad (2)$$

where Y_i is the density of specialized and innovative enterprises within the grid, X_{ij} represents factors influencing the spatial distribution of specialized and innovative enterprises, β_{bw_j} is the regression coefficient for the j -th influencing factor, subscript β_{bw_j} indicates the bandwidth applicable to the j -th influencing factor regression coefficient, and (u_i, v_i) denotes the centroid coordinates of the network.

3. Results

3.1. Spatial Distribution Characteristics of LGEs in the YRD and PRD

3.1.1. Overall Clustering Characteristics of LGEs in the YRD and PRD

(1) The overall distribution is characterized by high-value agglomeration, with the intensity of agglomeration in the PRD about twice that in the YRD.

Global G-statistic is used to analyze the spatial agglomeration characteristics of LGEs in the YRD and PRD. The results of the analysis (Table 3) show that at the 99% confidence level, the Global G-statistic indices of the observations in the two regions are significantly higher than the expected value of the random distribution, as shown by the Z-scores of the YRD (73.43) and PRD (44.55), which significantly exceed the critical value of the expected value of the random distribution, which is 2.58 ($p < 0.000001$). This indicates that the LGEs of the YRD and PRD show a significant high-value clustering pattern at the 99% confidence level.

Table 3. The global G-statistic for both regions (YRD and PRD).

Region Types	Z-Value	p-Value
YRD	73.43	0.000
PRD	44.55	0.000

Note: The Z-value indicates the number of standard deviations, and the p-value represents probability. Z and p are related such that when $Z < -2.58$ or $Z > +2.58$, $p < 0.01$, indicating a confidence level greater than 99%.

Ripley's K function was further used to explore the spatial distance scale differences in clustering, where the intensity of clustering was expressed as $L(d)$. The results show (Figure 4) that the observed value (ObservedK) is larger than the expected value (ExpectedK) at all scales and significantly higher than the high confidence interval (HiConfEnv). This indicates that at different distance scales, the LGEs in the YRD and PRD regions have obvious spatial agglomeration dynamics. In addition, the distance between the peak of the difference between the observed value of K and the expected value of K (DiffK) represents the peak of agglomeration intensity, which is about 65 km in the YRD and 30 km in the PRD, indicating that the spatial layout of enterprises in the YRD urban agglomeration is relatively dispersed and the homogenization of the industrial structure is obvious among the cities [59]. Enterprises need to seek resource synergies and produce agglomeration effects in a wider range, thus forming a peak agglomeration intensity of 65 km. Meanwhile, the integrated development plan for the YRD, which emphasizes coordinated regional development and resource sharing across the three provinces of Shanghai, Jiangsu, and Zhejiang, further promotes the clustering of businesses on a wider scale. In contrast, the PRD city cluster has well-developed intercity transportation and blurred administrative boundaries, the construction of a science and innovation industry system, synergy and openness, and infrastructure connectivity among cities [60]. Enterprises can access rich markets and resources in a smaller spatial area, which enhances their agglomeration intensity within 30 km.

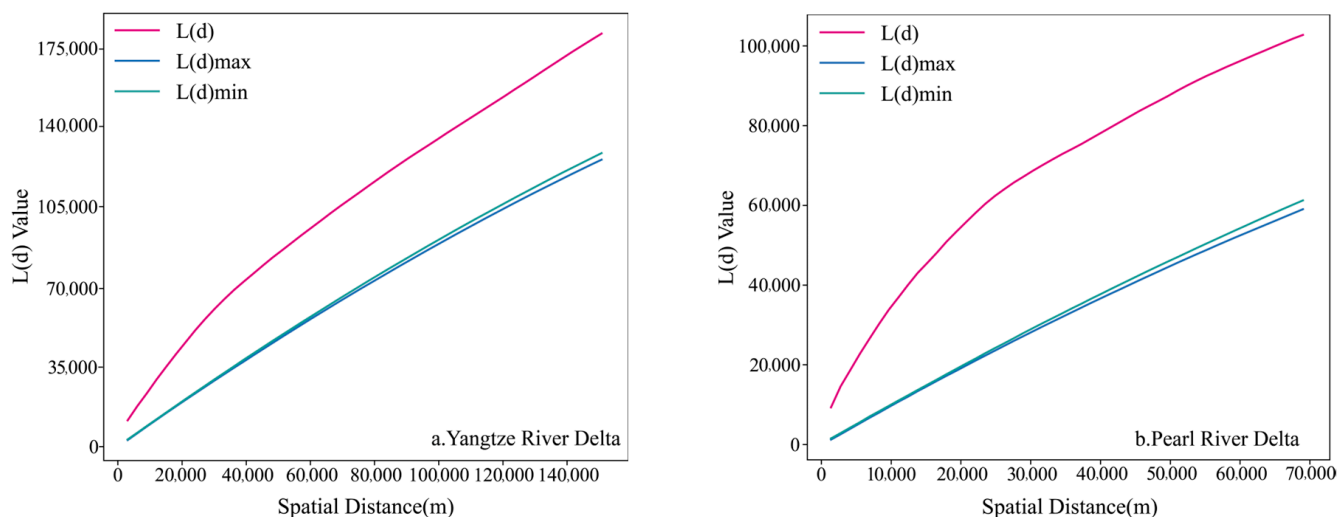


Figure 4. LGEs' Ripley's L(d) function analysis.

(2) The spatial distribution differences in kernel density and hot spot analysis between the YRD and PRD.

Combining the characteristics of L(d) curves, local clustering features of LGEs in the YRD (Figure 5) and PRD (Figure 6) regions are analyzed using kernel density and hot spot methods with radii of 65 km and 30 km, respectively.

Overall, the spatial distribution characteristics of LGEs in the YRD exhibit a “one-axis-three-core” pattern, with a relatively dispersed distribution (peak kernel density of 0.326). The “one-axis” refers to the east-west axis formed primarily by Shanghai and the “Suzhou-Wuxi-Changzhou” region in the downstream Yangtze River urban cluster. Shanghai serves as China's economic core city, financial center, and innovation hub. The “Suzhou-Wuxi-Changzhou” area, located in the core region of the YRD, is significantly influenced by Shanghai's driving effect and is a key manufacturing hub in China. The “three cores” are Hangzhou, Nanjing, and Ningbo along the river. In addition to geographic location advantages, Hangzhou and Nanjing, both provincial capitals and two of China's seven ancient capitals since ancient times, alongside Ningbo are regional economic centers. Meanwhile, Hangzhou has a well-developed e-commerce and internet economy and rich tourism resources. Nanjing has strong educational and scientific research strength and a deep historical and cultural heritage. Meanwhile, Ningbo, as a national-level marine economic demonstration zone, is one of the earliest coastal open cities in China, with the largest port in the world in terms of throughput. It can be seen that, combined with the characteristics of LGEs, Shanghai is the core of the development of the belt, and its driving role is obvious. Through the Yangtze River system, it connects “Suzhou, Wuxi and Changzhou” and Nanjing, as well as in the industrial docking and cooperation hubs of Hangzhou and Ningbo. Meanwhile, cold spots within the region exhibit a dispersed and low clustering distribution, primarily located in the western and eastern mountainous regions of Zhejiang Province (Tianmu Mountains, Tiantai Mountains), central areas of Jiangsu and Zhejiang provinces (such as Jurong, Gaoyou, and Taihu), and some coastal areas (such as Chongming and coastal regions of Nantong). These areas often consist of mountains, farmland, lakes, or regions distant from central urban areas.

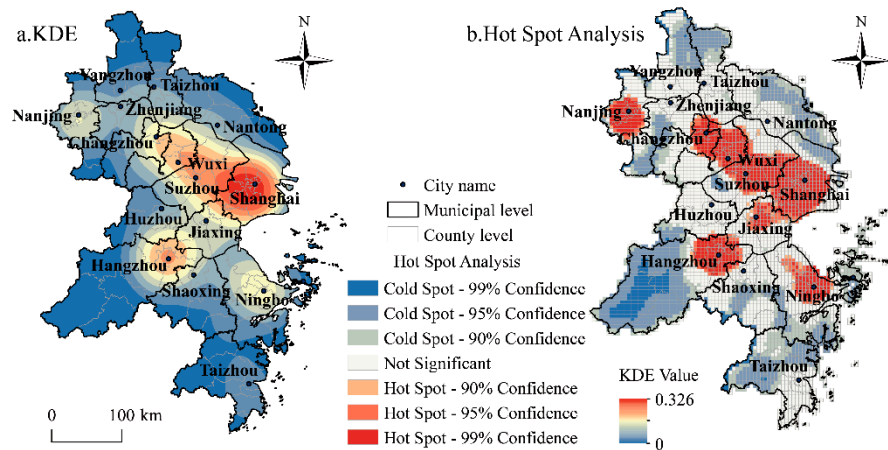


Figure 5. The kernel density distribution and hot spot analysis of LGEs in the YRD.

In the PRD region, LGEs exhibit an overall “single axial” spatial distribution pattern, characterized by prominent clustering features (peak kernel density of 1.301). This pattern forms along an urban axis linking Shenzhen, Dongguan, and Guangzhou. Shenzhen and Guangzhou serve as secondary core cities along this axis. Compared to the YRD, the PRD shows higher clustering intensity in its hot spots without other dispersed cores, which also contrasts with the differences observed in the peak clustering intensities (65 km for the YRD vs. 30 km for the PRD). Conversely, cold spots within the region exhibit a dispersed and low clustering pattern, primarily located in the northeast (such as Huidong County, Longmen County), northwest (such as Guangning County, Fengkai County), and southwest regions (such as Taishan City, Kaiping City). These areas consist mainly of mountains, farmland, forest parks, nature reserves, and regions distant from central urban areas.

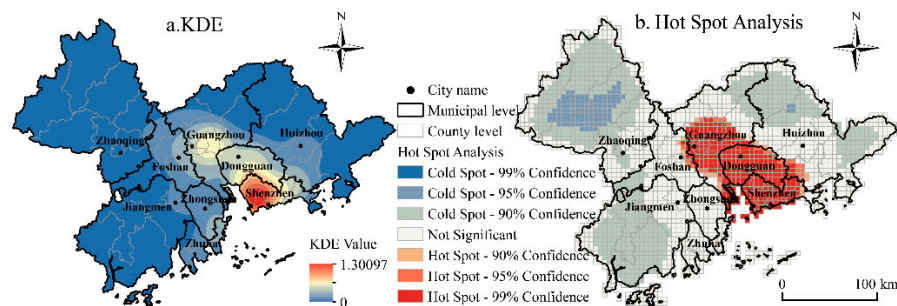


Figure 6. The kernel density distribution and hot spot analysis of LGEs in the PRD.

The distribution of LGEs in the YRD is more dispersed than in the PRD, probably due to the fact that the YRD region has a diversified economic structure due to the independent development of several cities as regional centers in the past, resulting in a more dispersed urban agglomeration. The PRD, conversely, has developed rapidly after the reform and opening up of China’s economy, with Shenzhen as a special economic zone leading to the formation of close ties and agglomeration of other cities in the region. While the traditional industries and extensive utilization of market resources in the YRD led to the distribution of firms over a wide area, the agglomeration effect of high-tech and export-oriented industries in the PRD enhanced the strong inter-city linkages.

3.1.2. Characteristics of the Spatial Distribution of LGEs in the YRD and PRD

(1) The YRD region focuses on traditional manufacturing industries, while the PRD region has a balanced development with high-tech industries at its core.

LGEs in the YRD and PRD are distributed in multiple industries, but there are significant differences in the proportion and concentration of their main industry types (Figure 7).

The distribution of the YRD's industry structure mainly focuses on the machinery and equipment manufacturing industry and the high-tech service industry, with the machinery and equipment manufacturing industry accounting for the highest proportion of 36.59%, reflecting the YRD region's deep technological foundation in traditional manufacturing. It also shows that the region has a well-developed industrial chain, which can provide integrated production and services from upstream supply of components to downstream manufacturing of complete machines. This was followed by the high-tech services industry (31.63%), which was higher than that in the PRD (26.08%). This reflects the YRD's comprehensive strengths in internet, software, and information technology services, research and experimental development, professional and technical services, and science and technology promotion and application services.

In contrast, the industry distribution of LGEs in the PRD region is more balanced. Among them, the high-tech manufacturing and high-tech service industries accounted for 25.20% and 26.08% respectively, showing the outstanding performance of the PRD in the high-tech sector. There is also a significant distribution of machinery and equipment manufacturing (19.22%), wholesale and retail trade (10.75%), other industries (11.96%), and food and textiles (4.77%), reflecting the region's diversified economic foundation. In summary, the PRD has a highly diversified economic structure. Its region covers not only traditional manufacturing and wholesale and retail industries but also high-tech manufacturing and service industries. Meanwhile, local governments in the PRD have long pursued diversified industrial policies to introduce high-end industries with policy support. They have also built autonomous innovation bases, as well as special industrial parks for fisheries, food, culture, and medical and health care, to achieve a balanced development of the cities in the Greater Bay Area [61].

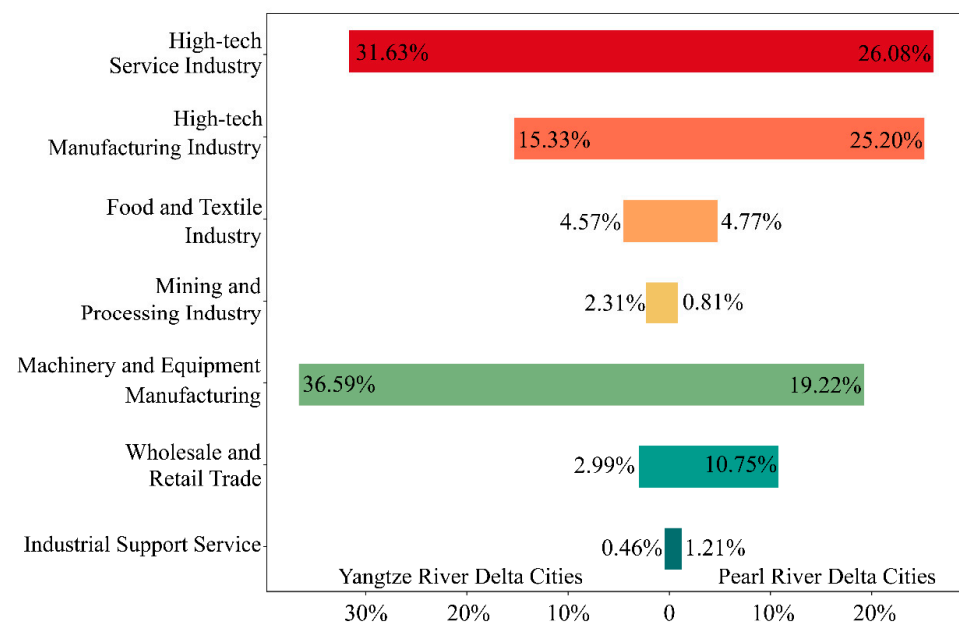


Figure 7. LGE industry distribution of YRD and PRD.

(2) There are obvious differences in the location and characteristics of the agglomeration of enterprises in different industries, but the overall characteristics of “multi-point agglomeration” are relatively stable.

In terms of overall layout, different industry cluster locations in the YRD are dominated by multiple patches and the cluster cities are more dispersed, while the PRD is dominated by point poles, and all are located in Guangzhou and Shenzhen (Figure 8). Among them, the layout of the high-tech service industry, mining and processing industry, food and textile industry, and industrial supporting services in the two regions is highly similar: the former are both clustered with dual cores, the latter are large-scale multi-area

clusters, and the latter two both spread outward from the pole to form a multi-patch distribution. The agglomeration characteristics of the machinery and equipment manufacturing, high-tech manufacturing, and wholesale and retail industries are quite different, but the characteristics in each region are similar, that is, they are multi-patch distribution in the YRD region, and point–pole distribution characteristics in the PRD region. The urban network in the YRD region has hierarchical and unbalanced characteristics. At the same time, there is a diversified division of functions among the central cities (such as finance and trade in Shanghai, e-commerce and high technology in Hangzhou, and manufacturing and foreign trade in Suzhou), so it presents a multi-patch distribution feature. The PRD region has largely formed a regional industrial division of labor with Shenzhen’s high-tech industries and Guangzhou’s service industries as the core. Therefore, enterprises are distributed in a point–polar distribution with Guangzhou and Shenzhen as the core.

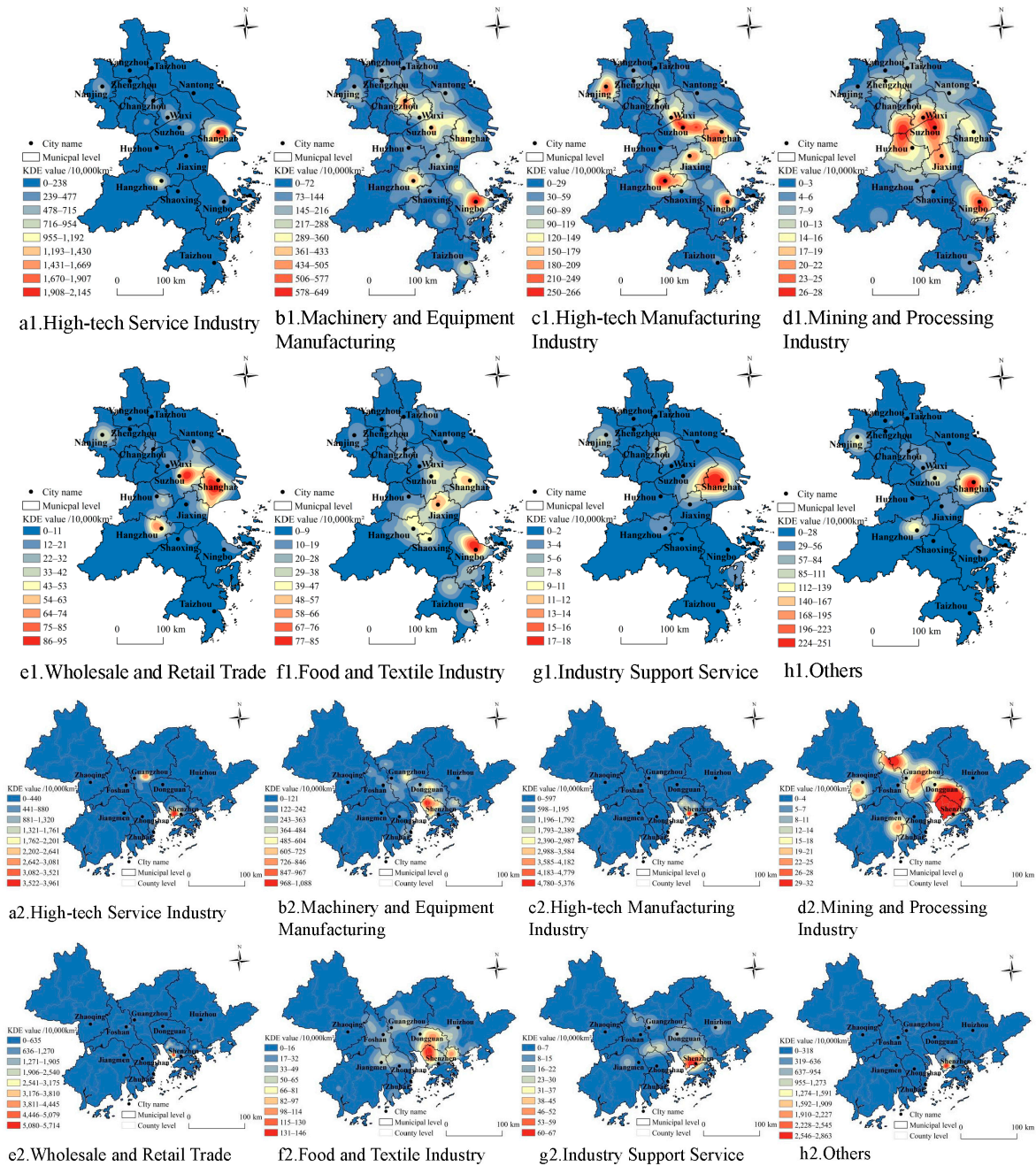


Figure 8. Kernel Density Distribution of LGEs by Industry.

3.2. Importance Analysis of Influencing Factors of LGEs Based on RF

In this paper, the contribution of several influences was assessed using the Random Forest (RF) method and ranked according to their relative importance (Figure 9). The training and validation data were validated by R^2 and SE (Table 4), which showed that high correlation coefficients ($R^2 > 0.9$) and low SE imply high model accuracy, and a low p -value (<0.05) indicates good model applicability.

Table 4. Results of the parameters used for the validation of model.

Region Types	Data Types	R^2	SE	p -Value
YRD	Training data	0.983	0.002	0.000
	Validation data	0.910	0.012	0.000
PRD	Training data	0.981	0.003	0.000
	Validation data	0.885	0.018	0.000

The research results indicate that among the top ten factors in both regions, industrial structure, logistics level, and industrial platforms are very important in both the PRD and the YRD. The significance of these factors is reflected in their widespread impact on enterprise development.

The industrial structure determines the types and scales of economic activities in a region, directly affecting the distribution of LGEs. Improved logistics levels can reduce costs and increase efficiency for business operations, thereby enhancing the competitiveness of enterprises. Industrial platforms provide concentrated resources and support for enterprises, including infrastructure, policy incentives, and supporting services, promoting enterprise growth and innovation.

However, there are differences in the impact of other factors between the YRD and the PRD. Housing prices and the labor market are among the top contributors in the YRD, while they rank lower in the PRD. This reflects the different economic structures and development models of the two regions. The YRD relies more on high-quality labor and downtown office space, making housing prices and the labor market significantly influential on enterprise location choices.

Enterprises in the PRD rely more on flexible industrial layouts and abundant migrant labor resources, making the impact of housing prices and the labor market relatively smaller. Conversely, instant and long-distance transport accessibility significantly contribute to the PRD, whereas their importance ranks lower in the YRD. This reflects different considerations for transportation infrastructure in enterprise location choices between the two regions. Enterprises in the PRD depend more on convenient intercity rail and highway systems to improve logistics efficiency and attract talent. Meanwhile, enterprises in the YRD rely more on rapid transport networks like aviation and high-speed rail to achieve technology transfer and efficient resource allocation.

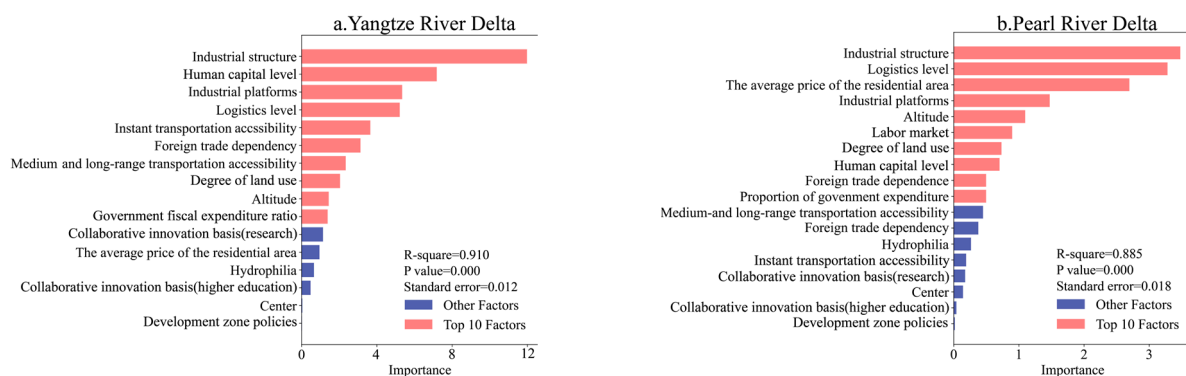


Figure 9. The relative importance of influencing factors in the YRD and PRD regions.

3.3. LGEs and Spatial Heterogeneity of Influencing Factors

In studying the factors influencing the agglomeration of LGEs in the YRD and the PRD, the top ten key factors in the two regions that together account for the top 90% of the explanatory power in the RF model were selected for analysis. The selection of these factors ensured the accuracy and reliability of the analysis while avoiding model complexity. The MGWR model analyses (Tables 5 and 6) were able to reveal the direction and strength of the factors on the agglomeration of LGEs, as well as the spatial heterogeneity of the response of these factors to enterprise (Figures 10 and 11), providing statistically significant drivers.

Table 5. Summary Statistics for MGWR Coefficient Estimates of LGEs in the YRD.

Impact Dimension	Variable	Bandwidth (% of Extent)	Significance (% of Features)	Coefficient			
				Mean	Std.Dev.	Min	Max
Natural geography and location	Altitude	53.15 (7.21)	896 (18.01)	−0.6163	1.5723	−7.1835	2.9775
	Degree of land use	53.15 (7.21)	1736 (34.89)	0.0048	0.0432	−0.1312	0.102
Land use and cost	Average listing price of the community	53.15 (7.21)	2026 (40.72)	0.062	0.0822	−0.1572	0.3113
	Human capital level	53.15 (7.21)	3520 (70.75)	0.0656	0.2519	−0.7997	1.6185
Industrial development basis	Industrial platforms	53.15 (7.21)	1975 (39.70)	0.1024	0.1505	−0.0472	0.59
	Industrial structure	53.15 (7.21)	2482 (49.89)	−0.0176	0.0482	−0.1717	0.1552
	Labor market	737.05 (100)	4975 (100.00)	0.0113	0.0001	0.0111	0.0114
External supporting condition	logistics level	53.15 (7.21)	4416 (88.76)	0.1877	0.1543	−0.1989	0.5953
	Government fiscal expenditure ratio	53.15 (7.21)	3004 (60.38)	−0.0404	0.1037	−0.44	0.1827
	Foreign trade dependency	53.15 (7.21)	3982 (80.04)	0.0992	0.0861	−0.4189	0.2924

Table 6. Summary Statistics for MGWR Coefficient Estimates of LGEs in the PRD.

Impact Dimension	Variable	Bandwidth (% of Extent)	Significance (% of Features)	Coefficient			
				Mean	Std.Dev.	Min	Max
Natural geography and location	Altitude	50.25 (9.70)	357 (14.23)	0.0195	0.0575	−0.3114	0.222
Transportation accessibility	Instant transportation accessibility	202.847 (39.16)	31 (1.24)	−0.0043	0.0045	−0.0224	0.0024
	Medium and long-distance transportation accessibility	50.25 (9.70)	927 (36.96)	−0.1288	0.1986	−1.1846	0.0141
Land use and cost	Degree of land use	518 (100.00)	0 (0.00)	−0.0051	0.0001	−0.0052	−0.0049
Scientific research and innovation conditions	Human capital level	50.25 (9.70)	1151 (45.89)	0.0129	0.1295	−0.2892	0.301
Industrial development basis	Industrial platforms	104.73 (20.22)	2194 (87.48)	0.0821	0.0309	0.0273	0.1448
	Industrial structure	50.25 (9.70)	771 (30.74)	0.0496	0.1396	−0.1058	0.7354
External supporting condition	Logistics level	50.25 (9.70)	2508 (100.00)	0.2405	0.1372	0.0602	0.713
	Government fiscal expenditure ratio	50.25 (9.70)	1133 (45.18)	−0.0255	0.1524	−0.7095	0.3453
	Foreign trade dependency	50.25 (9.70)	1621 (64.63)	0.0956	0.0862	0.1064	0.2303

(1) Natural geographic and location

There are significant but different effects of physical geographic location factors on the clustering of LGEs in the YRD and PRD regions. In the YRD region, altitude plays a negative role in the regions of Suzhou, Wuxi, Changzhou, Jiaxing, southeast of Shaoxing, and northwest of Hangzhou, mainly because higher altitude areas are usually far away from the economic centers and major markets, which increases the difficulty of sales and services

and limits the ability of firms to expand in the market. In contrast, in the PRD, especially in Foshan, Zhongshan, and Jiangmen, altitude has a positive effect on the concentration of LGEs. The cities of Guangzhou, Foshan, Zhongshan, Jiangmen, and Dongguan in the PRD have low altitudes (5–20 m) and the overall terrain is flat. The PRD is crisscrossed by rivers, and low-altitude areas are mostly depressions, which are not conducive to business operations, and LGEs are more inclined to cluster in slightly higher-altitude areas.

(2) Industrial development basis

Industrial structural factors exhibit significant but different impacts on the clustering of LGEs in the YRD and the PRD. In the YRD, particularly in cities like Nanjing, Hangzhou, and Ningbo, the industrial structure has a negative influence on the clustering of LGEs. This is mainly due to a high proportion of traditional industries in the industrial structure of these areas, which limits the access of enterprises to technological and market resources, thereby hindering their clustering [62]. In the PRD region, the industrial structure in Shenzhen and its surrounding urban circle (Guangzhou, Huizhou, Zhuhai, and parts of Zhongshan) has a positive relationship with the concentration of LGEs. This is mainly due to the fact that the PRD region, especially Shenzhen and Guangzhou, has a high proportion of high-tech industries and modern service industries [63], which provide a wealth of innovation resources and co-operation opportunities for LGEs. In addition, the PRD region has a high degree of market openness and a strong degree of internationalization, which provides a broader market and development space for LGEs.

In the YRD and PRD regions, industrial platforms (the number of industrial parks) show a positive impact on the agglomeration of LGEs. The clustering of industrial parks plays a key role in the innovation and growth of enterprises. For example, Guangzhou, Shenzhen, and Foshan have a large number of industrial parks, and high-tech industrial parks such as Caohejing, Songjiang Economic and Technological Development Zone and G60 Science and Innovation Corridor in south-west Shanghai, and Zhoushan Lingang Industrial parks are all important industrial platforms. The agglomeration of industrial parks can enhance regional competitiveness and promote knowledge exchange and technology transfer among enterprises [64]. In addition, industrial parks have a high degree of specialization and supporting service levels. In addition to perfect infrastructure, they also provide various support services such as research and development support, financing services, and market promotion. These factors provide important support for the rapid development of LGEs [65].

There is no significant spatial difference in the effect of population density in the YRD region. Overall, there is a significant positive relationship between population density and the agglomeration of LGEs in the YRD region. This suggests that high population density areas provide important support for the development of LGEs, especially in terms of the labor market. By providing a rich labor supply and a diverse pool of talent, high-population-density regions enhance the ability of enterprises to acquire the necessary skills and human capital to meet their expansion and operational requirements. In addition, the diversity of talent in high-population-density regions enables firms to optimize the allocation of human resources and promote organizational competitiveness and innovation.

(3) External supporting conditions

The level of logistics shows a significant positive effect on the agglomeration of LGEs in the YRD and PRD regions. The logistics level in the YRD is a local variable, as evidenced by the strong positive and significant influence in the urban areas of Shanghai, Suzhou, and Nanjing, as well as in the mountainous areas in the south-west of the country. The logistics level in the PRD is a local variable, with a significant positive effect on the agglomeration of LGEs in the whole region, with the intensity of the effect being higher in Shenzhen, Zhuhai, and parts of Zhongshan and Dongguan. The analysis of the possible reasons for this is that in urban areas, such as Shanghai and Shenzhen, advanced logistics networks and facilities (including highways, ports, and logistics parks) reduce operating and transport costs and enhance the operational efficiency and market competitiveness of enterprises.

The construction of excellent highway networks further reduces transport costs, making regions with more developed logistics infrastructure more attractive to the clustering of LGEs [66]. In mountainous areas, improvements in the level of infrastructure and logistics significantly increase the concentration of firms, and in areas of logistical bottlenecks, any upgrading generates greater marginal effects, significantly increasing the attractiveness and agglomeration effect for enterprises.

Government finance exhibits a negative effect on the agglomeration of LGEs in the YRD and PRD. Government finance in the YRD is a local variable, and its spatial impact is particularly significant in Suzhou, Wuxi, Changzhou, Nantong, Ningbo, Hangzhou, and their neighboring cities. Government fiscal expenditure in the PRD is a local variable, showing negative impacts in Shenzhen, Guangzhou, Dongguan, Foshan, Zhongshan, and Zhuhai, with the intensity of the negative impacts being relatively high in Shenzhen. The possible reasons for this are analyzed as follows: firstly, the resource crowding-out effect leads to an unbalanced distribution of financial resources, and a high proportion of public budgetary expenditure may lead to a reduction in the government's direct support for enterprises, such as funds for innovation and enterprise subsidies, which affects the development of LGEs. Secondly, a high proportion of government fiscal expenditure may lead to excessive market intervention, inhibiting the effective operation of the market mechanism and affecting the independent innovation and market competitiveness of enterprises.

Foreign trade dependency has a positive impact on the agglomeration of LGEs in the YRD and PRD. The foreign trade dependency of the YRD is a local variable, and its spatial impact is particularly significant in Nanjing, Suzhou, Nantong, Ningbo, Hangzhou, and their surrounding cities. Foreign trade dependency in the PRD is a localized variable, showing a positive impact in Shenzhen, Guangzhou, Foshan, Zhaoqing, and parts of Dongguan. These regions have well-developed foreign trade service systems, including customs, international logistics, and trade finance services, as well as strong manufacturing industries, which enable enterprises to gain better access to international markets, expand overseas business, and enhance market competitiveness. In addition, these regions have improved the productivity and market responsiveness of their enterprises by optimizing the allocation of resources and strengthening international cooperation.

(4) Scientific research and innovation conditions

In the YRD and PRD regions, the level of human capital shows a significant positive effect on the agglomeration of LGEs. The level of human capital in the YRD is a local variable (mean 0.0656) with a significant positive effect in Hangzhou and its neighboring cities of Huzhou, Jiaxing, Shaoxing, and Taizhou. These cities are home to a large number of higher education institutions, have a relatively high number of students enrolled in them as a percentage of the regional population, and are located in close proximity to each other. Geographical proximity creates a good flow of talent and interaction, with high-quality talent not only serving local companies but also flowing to neighboring companies, facilitating the clustering and development of LGEs in the region. The level of human capital in the PRD is a local variable (mean 0.0129) with a significant positive effect in parts of Jiangmen. The industrial structure of Jiangmen is relatively traditional and is in a stage of transformation and upgrading, where high levels of human capital are particularly important for promoting innovation and the growth of LGEs. The high-quality labor force promotes industrial upgrading and technological innovation, thereby attracting and promoting the agglomeration of LGEs.

(5) Land Use and Cost

The degree of land use in the YRD is a local variable (mean 0.0048), which mainly has a significant positive effect in the Suzhou, Wuxi, Changzhou, and Hangzhou regions. The most likely reason for this is the high proportion of urban construction land in these areas, especially in economically developed areas such as Suzhou Industrial Park and High-Tech Zone, and Hangzhou's Xihu and Binjiang Districts. A high proportion of built-up land in towns and cities implies a high degree of urbanization and good infrastructure, making it a

fertile ground for LGEs to cluster. The effect of land use in the PRD on the agglomeration of LGEs is spatially insignificant and not significantly different.

The average listing price of cells in the YRD as a local variable has a significant positive effect on the agglomeration of LGEs (mean 0.0992), and its effect is spatially significant, especially in the cities of Shanghai, Nanjing, Hangzhou, Suzhou, Wuxi, Ningbo, and Nantong. High-price areas usually reflect strong economic dynamism and market demand [67], and these areas have well-developed infrastructure and public services, including transport, education, and healthcare, which provide a favorable supportive environment for business operations. In addition, areas with high property prices are usually also areas with high spending power, which provides a broad market opportunity for LGEs and promotes the agglomeration and development of businesses.

(6) Transportation accessibility

The medium- and long-distance transport accessibility in the PRD is a local variable (mean -0.1288) with significant negative impacts in Shenzhen and the surrounding urban areas (Guangzhou, Foshan, Zhongshan, Jiangmen, Zhuhai, Dongguan, Huizhou, and parts of Shenzhen), with higher intensities in Dongguan, Guangzhou, and parts of Shenzhen. Analyzing the possible reasons shows the following. (1) Cost and resource layout: these areas may be more suitable as logistics and transport transit points than as long-term business locations. Businesses prefer locations close to markets, raw materials, and labor to optimize supply chains and reduce costs. (2) Comprehensive support and business needs: These areas are mostly planned for transport infrastructure rather than commercial or industrial development. This results in limited land that can be used for corporate offices or production, and a lack of well-developed commercial, educational, medical, and other amenities, which further reduces the chances of companies gathering here. Immediate transport accessibility in the PRD is a local variable and is insignificant in most areas except for a small number of areas in eastern Huizhou where it is significantly negative.

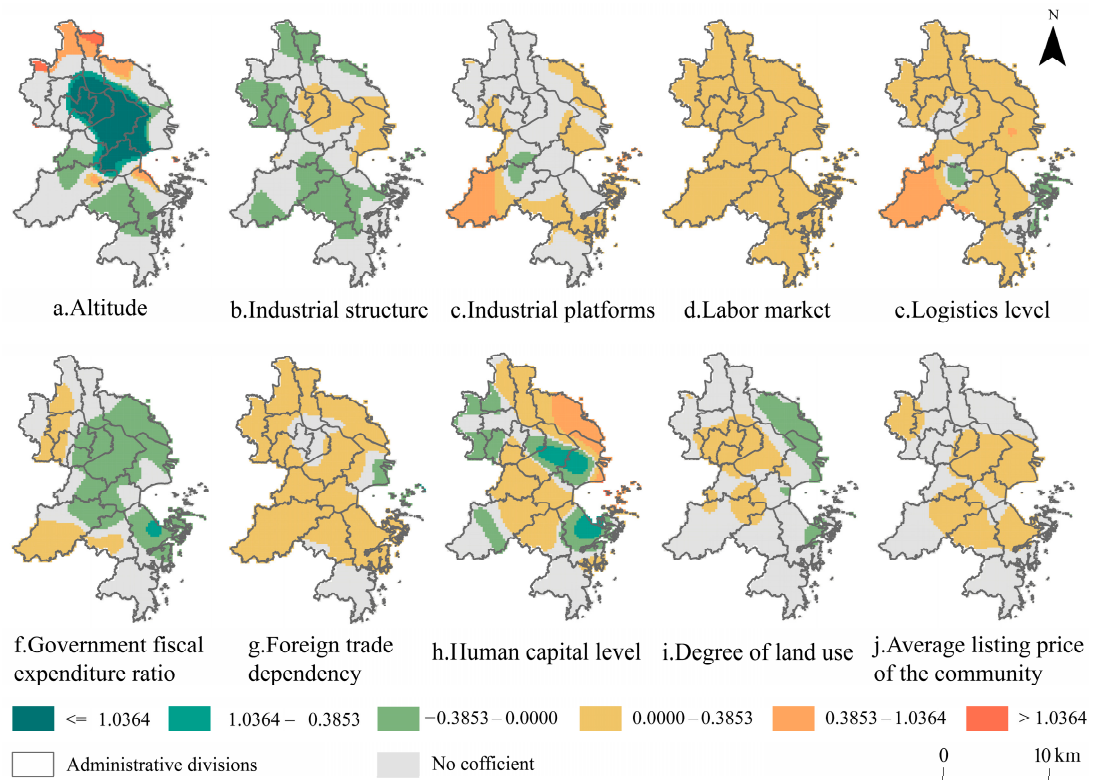


Figure 10. Patterns of spatial differentiation in factors influencing the distribution of LGEs in the YRD.

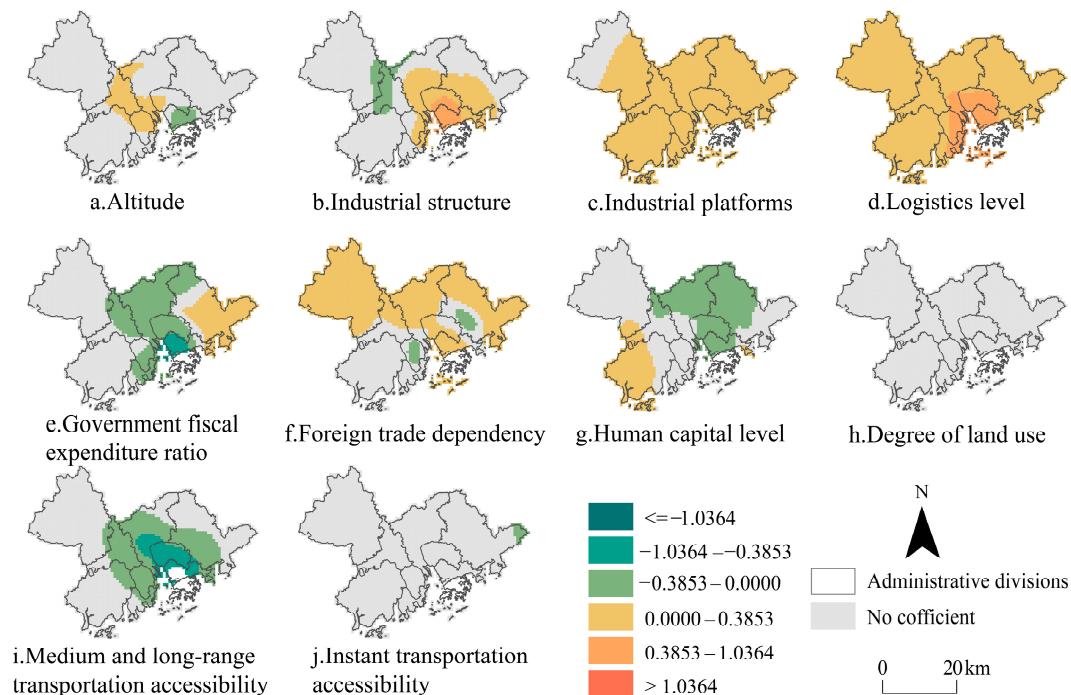


Figure 11. Patterns of spatial differentiation in factors influencing the distribution of LGEs in the PRD.

4. Discussion and Conclusions

4.1. Discussion

In terms of industrial distribution characteristics, the results of this paper show that LGEs in the YRD are still dominated by traditional manufacturing, which is consistent with the conclusion of Xu et al. that there are a large number of traditional manufacturing enterprises in the YRD [68]. This further shows that the YRD needs to eliminate traditional enterprises with high pollution and backward technology and accelerate the development of modern service industries and strategic emerging industries. In the PRD region, this paper finds that LGEs account for the highest proportion of high-tech manufacturing and high-tech services, which is consistent with the conclusion of previous studies such as Li et al. that the PRD region has transformed from a labor-intensive world factory to a global urban region driven by technological innovation [69].

In terms of the main factors affecting enterprise agglomeration, the results of this paper are consistent with the previous research conclusions that industrial structure, industrial parks, logistics, and other factors have a significant impact on enterprise agglomeration. Based on a large number of previous studies [13,16,17], this paper starts from the complexity and heterogeneity of the two mega-city agglomerations of the YRD and the PRD and comprehensively considers seven factors, namely, natural geography and location [27], transportation accessibility [28], land use and cost [28,30], living convenience [36,37], scientific research and innovation conditions [31,32], industrial development foundation [33], and external support conditions [34,35], and attempts to explain the role of different dimensional influencing factors in the distribution of small giant enterprises. As the regions with the highest degree of openness and the most active economy in the country, the YRD and the PRD are undoubtedly excellent objects for studying the cultivation environment and growth laws of specialized, refined, and innovative enterprises. The conclusions of this paper can provide reference and inspiration for other regions.

This paper provides a new perspective for understanding the spatial agglomeration characteristics and influencing factors of LGEs in the YRD and the PRD. By analyzing the spatial distribution characteristics and main influencing factors of enterprises in the two regions, this study not only reveals the common and different reasons for the agglomeration of LGEs in different urban agglomerations but also clarifies the “geographical boundaries” of

different influencing factors. These research results have important reference value for the cultivation of LGEs and the formulation of regional economic development policies.

This study has the following possible limitations: First, the data on LGEs used in this paper are cross-sectional data, and the changes in enterprises in the time dimension cannot be observed. Future research can use cross-year data to explore the evolution of LGEs in the dual dimensions of time and space to provide a more comprehensive dynamic analysis perspective. Secondly, this paper selected $5\text{ km} \times 5\text{ km}$ as the grid scale. Although this can reflect the local agglomeration characteristics of enterprises, it may also ignore the regional linkage effects or micro-agglomeration characteristics that may be included in other scale ranges. Therefore, future research can try different grid scales for comparison, such as $1\text{ km} \times 1\text{ km}$ or $10\text{ km} \times 10\text{ km}$, so as to provide a more multi-dimensional analysis perspective for the spatial layout of enterprises. Finally, this paper takes the grid in the study area as the research object, ignoring the observation of the association between individual enterprises and the external relationship network in which the enterprises are located. In the future, network research methods and paradigms can be applied, such as collaboration between large, medium, and small enterprises, industry-university-research cooperation, and urban network associations, to further refine the research related to specialization, refinement, and innovation.

4.2. Conclusions

This study selects LGEs in the YRD and PRD regions as research subjects, integrating geographical spatial data with socioeconomic statistics. Through spatial autocorrelation analysis, multi-distance spatial clustering analysis, kernel density analysis, and other methods, it investigates the spatial distribution patterns and characteristics of these enterprises. Additionally, RF and MGWR are used to explore the importance of factors influencing the spatial distribution of LGEs at a $5\text{ km} \times 5\text{ km}$ grid scale, as well as their spatial heterogeneity. The main conclusions are as follows:

- (1) In terms of spatial distribution characteristics, LGEs show significant spatial agglomeration in the YRD and the PRD. The peak of agglomeration in the YRD occurs at 65 km, forming a “one-axis-three-core” distribution pattern centered around Shanghai and the “Suzhou-Wuxi-Changzhou” area, with Nanjing, Hangzhou, and Ningbo as core cities. In contrast, in the PRD, the clustering peak occurs at 30 km, characterized by a “single axial” distribution pattern along the line from Shenzhen to Guangzhou. The clustering intensity in the PRD is relatively higher compared to the YRD.
- (2) In terms of industrial distribution characteristics, the YRD is dominated by traditional manufacturing industries, supplemented by high-tech service industries, both sectors being prominent. In contrast, the PRD has a balanced development of high-tech manufacturing and service industries. The clustering locations and characteristics of enterprises in different industries exhibit some variations, but overall, they show a “multi-cluster” feature. The YRD is characterized by multi-patch distribution, while the PRD is characterized by point-polar distribution.
- (3) Regarding the main factors influencing the clustering of LGEs in the YRD and PRD, their spatial distribution is influenced by similar factors. These factors primarily include industrial structure, industrial platforms, logistics level, proportion of government fiscal expenditure, dependence on foreign trade, human capital level, and altitude. Among these, industrial structure, industrial platforms, and logistics level exert the greatest influence. In the YRD, the presence of multiple cores is significant, with a greater emphasis on land use costs and human capital. Conversely, in the PRD, there is a stronger focus on transportation accessibility.
- (4) There are scale effect differences in the role of factors influencing the spatial distribution of LGEs in the YRD and PRD regions. Among the seven factors that have a significant impact on the agglomeration of LGEs, industrial platforms, logistics level, foreign trade dependence, and human capital level all have a positive impact, while government financial expenditure has a negative impact. Although the impact

direction of industrial structure is opposite in the two regions, its overall impact pattern remains consistent. The positive or negative impact of natural geographical location differs between the two regions, but it is not a primary factor.

In response to the above conclusion, this paper puts forward the following suggestions: (1) This study suggests that local governments should optimize the spatial layout of emerging LGEs based on the discovery of the spatial distribution characteristics of two places, thus giving full play to the core role of the “one axis, three cores” and “single axis” central axis. (2) This study found that the industrial structure of the YRD is still dominated by manufacturing. Resource-intensive and environmentally polluting industries often lack economic resilience and are unsustainable at the ecological level. Regions need to improve their competitiveness through the transformation of industrial structure, especially in high-tech, high-value-added industries, while developing green technologies and industries to promote the transformation of industrial structure in a green, low-carbon, and sustainable direction. (3) This article points out that factors such as industrial parks, government policies, and human capital have a significant impact on enterprise agglomeration. The government should improve the construction of industrial parks, provide appropriate preferential policies in optimizing the layout, promote high-potential SMEs to integrate capital, talents, and other factors through specialized markets and competitive industrial chains, and attract enterprises to develop professionally and form LGEs.

Author Contributions: Conceptualization, J.D. and G.C.; methodology, J.D., F.Y.; software, J.D.; validation, Z.Z., Y.X. and X.Y.; formal analysis, J.D.; investigation, Y.X.; resources, G.C.; data curation, J.D.; writing—original draft preparation, J.D.; writing—review and editing, Z.Z., F.Y.; visualization, Z.Z.; supervision, G.C.; project administration, X.Y.; funding acquisition, G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 42071172.

Data Availability Statement: Data are available upon request.

Acknowledgments: We are deeply grateful to the reviewers for their insightful comments and suggestions, which have significantly improved the manuscript. We would also like to extend our heartfelt thanks to Ziyi Jia from the University of Cambridge for her invaluable advice on the overall experimental design, and to Zixiao Hu for his assistance with data processing. Additionally, we appreciate the support from the project on Spatial and Temporal Knowledge Mapping and Thematic Mapping of Jiangnan Water Town Literary Lineage.

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

Appendix A

Table A1. Industry Classification of LGEs.

Industry Classification	Specific Industry Names and Codes
High-tech Service Industry	Internet Services (64); Software and IT Services (65); Research and Experimental Development (73); Professional Technical Services (74); Technology Transfer and Application Services (75)
High-tech Manufacturing Industry	Chemical Raw Materials and Chemical Products Manufacturing (26); Pharmaceutical Manufacturing (27); Chemical Fiber Manufacturing (28); Computer, Communication, and Other Electronic Equipment Manufacturing (39); Instrument and Apparatus Manufacturing (40); Ecological Protection and Environmental Governance (77)

Table A1. Cont.

Industry Classification	Specific Industry Names and Codes
Food and Textile Industry	Processing of Food from Agricultural Products (13); Food Manufacturing (14); Manufacture of Beverages, Alcoholic Drinks and Refined Tea (15); Tobacco Products Industry (16); Textile Industry (17); Manufacture of Textile Wearing Apparel, and Accessories (18); Leather, Fur, Feather (Plume), and Related Products, and Footwear Manufacturing (19); Processing of Wood and Manufacture of Products of Wood, Bamboo, Rattan, Palm, and Straw (20); Furniture Manufacturing (21); Paper and Paper Products Industry (22); Printing and Reproduction of Recorded Media (23); Manufacture of Articles for Culture, Education, Arts and Crafts, Sports and Entertainment (24); Rubber and Plastics Products Industry (29); Other Manufacturing (41)
Mining and Processing Industry	Nonferrous Metal Mining and Dressing (09); Non-metallic Mineral Mining and Dressing (10); Petroleum, Coal, and Other Fuel Processing Industry; Non-metallic Mineral Products Industry (25); Ferrous Metal Smelting and Rolling Processing Industry (31); Nonferrous Metal Smelting and Rolling Processing Industry (32)
Machinery and Equipment Manufacturing	General Equipment Manufacturing (34); Special Equipment Manufacturing (35); Automobile Manufacturing (36); Manufacture of Railways, Ships, Aerospace, and Other Transport Equipment (37); Electrical Machinery and Equipment Manufacturing (38); Comprehensive Utilization of Waste Resources (42); Repair of Metal Products, Machinery, and Equipment (43); Metal Products Industry (33)
Wholesale and Retail Trade	Wholesale Trade (51); Retail Trade (52)
Industrial Support Service	Electricity and Heat Production and Supply (44); Residential Building Construction (47); Civil Engineering Construction (48); Building Installation Services (49); Road Transport (54); Capital Market Services (67); Leasing Services (71); Business Services (72)
Others	Agriculture (01); Livestock Farming (03); Other Mining Industries (12); Water Production and Supply (46); Building Decoration and Other Construction (50); Multimodal Transport and Transport Agency Services (58); Telecommunications and Satellite Transmission Services (63); Real Estate (70); Water Management (76); Public Facility Management (78); Land Management (79); Residential Services (80); Vehicle, Electronics, and Consumer Goods Repair (81); Other Services (82); Health Services (84)

Note: The numbers in parentheses are the GB/T 4754-2017 National Economic Industry Classification codes.

References

- Simon, H. Lessons from Germany's midsize giants. *Harv. Bus. Rev.* **1992**, *70*, 115–121.
- Simon, H. *Hidden Champions in the Chinese Century*; Springer Books; Springer: Berlin/Heidelberg, Germany, 2022.
- Garcia-Vega, M. Does technological diversification promote innovation?: An empirical analysis for European firms. *Res. Policy* **2006**, *35*, 230–246. [[CrossRef](#)]
- Ellison, G.; Glaeser, E.L.; Kerr, W.R. What causes industry agglomeration? Evidence from coagglomeration patterns. *Am. Econ. Rev.* **2010**, *100*, 1195–1213. [[CrossRef](#)]
- Söderbom, M.; Weng, Q. Multi-product firms, product mix changes and upgrading: Evidence from China's state-owned forest areas. *China Econ. Rev.* **2012**, *23*, 801–818. [[CrossRef](#)]
- Shao, Y. Research on the Impact of the Specialized, Refined, Unique, and Innovative "Little Giant" Policy on the Small and Medium Enterprises' Innovation. *J. Innov. Dev.* **2024**, *6*, 12–19. [[CrossRef](#)]
- Simon, H. Target Market China. In *Hidden Champions in the Chinese Century: Ascent and Transformation*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 99–112.
- Lei, L.; Wu, X.; Tan, Z. The growth of hidden champions in China: A cognitive explanation from integrated view. *Chin. Manag. Stud.* **2020**, *14*, 613–637. [[CrossRef](#)]
- Li, N.; Song, S. A quasi-natural experimental study on enterprise innovation driven by urban agglomeration policies in China. *Sci. Rep.* **2023**, *13*, 10297. [[CrossRef](#)]
- Jang, S.; Kim, J.; von Zedtwitz, M. The importance of spatial agglomeration in product innovation: A microgeography perspective. *J. Bus. Res.* **2017**, *78*, 143–154. [[CrossRef](#)]
- Porter, M.E. *The Competitive Advantage of Nations*; Macmillan: London, UK, 1990.

12. Krugman, P. Increasing returns and economic geography. *J. Bus. Res.* **1991**, *99*, 483–499. [[CrossRef](#)]
13. Asheim, B.T. Industrial districts as “learning regions”: A condition for prosperity. *Eur. Plan. Stud.* **1996**, *4*, 379–400. [[CrossRef](#)]
14. Gertler, M.S. Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *J. Econ. Geogr.* **2003**, *3*, 75–99. [[CrossRef](#)]
15. Giuliani, E. The selective nature of knowledge networks in clusters: Evidence from the wine industry. *J. Econ. Geogr.* **2007**, *7*, 139–168. [[CrossRef](#)]
16. Ellison, G.; Glaeser, E.L. Geographic concentration in US manufacturing industries: A dartboard approach. *J. Polit. Econ.* **1997**, *105*, 889–927. [[CrossRef](#)]
17. Ravix, J.-L. Localization, innovation and entrepreneurship: An appraisal of the analytical impact of Marshall’s notion of industrial atmosphere. *J. Innov. Econ. Manag.* **2014**, 63–81. [[CrossRef](#)]
18. Moore, S.B.; Manring, S.L. Strategy development in small and medium sized enterprises for sustainability and increased value creation. *J. Clean. Prod.* **2009**, *17*, 276–282. [[CrossRef](#)]
19. Kuivalainen, O.; Sundqvist, S.; Saarenketo, S.; McNaughton, R. Internationalization patterns of small and medium-sized enterprises. *Int. Mark. Rev.* **2012**, *29*, 448–465. [[CrossRef](#)]
20. Das, M.; Rangarajan, K.; Dutta, G. Corporate sustainability in small and medium-sized enterprises: A literature analysis and road ahead. *J. Indian Bus. Res.* **2020**, *12*, 271–300. [[CrossRef](#)]
21. Zhu, H.; Liu, R.; Chen, B. The Rise of Specialized and Innovative Little Giant Enterprises under China’s ‘Dual Circulation’ Development Pattern: An Analysis of Spatial Patterns and Determinants. *Land* **2023**, *12*, 259. [[CrossRef](#)]
22. Tang, G.; Wang, L.; Zheng, T.; Wu, W. What types of business environment fosters the emergence of more specialized and sophisticated “little giant” enterprises?—An empirical study based on the TOE framework and configuration adaptation theory. *Manag. Decis. Econ.* **2024**, *45*, 1557–1572. [[CrossRef](#)]
23. Jianjun, D.; Xian, L.; Diankun, W.; Jinwen, Y. Spatial Distribution and Influencing Factors of China’s National-level “Little Giant” Enterprises. *Econ. Geogr.* **2022**, *42*, 109–118.
24. Li, J. The Situation and Outlet of the Development of China’s “Little Giant” Enterprises. *Reform* **2021**, *10*, 101–113.
25. Li, L.; Zhang, X. Spatial evolution and critical factors of urban innovation: Evidence from Shanghai, China. *Sustainability* **2020**, *12*, 938. [[CrossRef](#)]
26. Wu, W.; Ding, Z.; Huang, K.; Song, Y.; Dong, H. Spatial distribution of enterprise communities and its implications based on POI data: Case of Xi’an, China. *J. Urban Plan. Dev.* **2021**, *147*, 05021028. [[CrossRef](#)]
27. Rasvanis, E.; Tselios, V. Do geography and institutions affect entrepreneurs’ future business plans? Insights from Greece. *J. Innov. Entrep.* **2023**, *12*, 3. [[CrossRef](#)]
28. Iseki, H.; Eom, H. Impacts of rail transit accessibility on firm spatial distribution: Case study in the metropolitan area of Washington, DC. *Transp. Res. Record.* **2019**, *2673*, 220–232. [[CrossRef](#)]
29. Xu, Z.; Huang, J.; Jiang, F. Subsidy competition, industrial land price distortions and overinvestment: Empirical evidence from China’s manufacturing enterprises. *Appl. Econ.* **2017**, *49*, 4851–4870. [[CrossRef](#)]
30. Wang, L.; Yang, Y. Political connections in the land market: Evidence from China’s state-owned enterprises. *Real Estate Econ.* **2021**, *49*, 7–35. [[CrossRef](#)]
31. Zheng, S.; Du, R. How does urban agglomeration integration promote entrepreneurship in China? Evidence from regional human capital spillovers and market integration. *Cities* **2020**, *97*, 102529. [[CrossRef](#)]
32. García-Estévez, J.; Duch-Brown, N. The relationship between new universities and new firms: Evidence from a quasi-natural experiment in Spain. *Reg. Stud. Reg. Sci.* **2020**, *7*, 244–266. [[CrossRef](#)]
33. Novyidarskova, E. The Effectiveness of Industrial Parks in the Regional Economy. *Probl. Econ. Transit.* **2020**, *62*, 617–621. [[CrossRef](#)]
34. Wen, H.; Lee, C.-C.; Zhou, F. How does fiscal policy uncertainty affect corporate innovation investment? Evidence from China’s new energy industry. *Energy Econ.* **2022**, *105*, 105767. [[CrossRef](#)]
35. Wang, S.; Ahmad, F.; Li, Y.; Abid, N.; Chandio, A.A.; Rehman, A. The Impact of Industrial Subsidies and Enterprise Innovation on Enterprise Performance: Evidence from Listed Chinese Manufacturing Companies. *Sustainability* **2022**, *14*, 4520. [[CrossRef](#)]
36. Litman, T. *Comprehensive Parking Supply, Cost and Pricing Analysis*; Victoria Transport Policy Institute: Victoria, BC, Canada, 2022.
37. Wang, B.; Wen, B. The spatial distribution of businesses and neighborhoods: What industries match or mismatch what neighborhoods? *Habitat Int.* **2021**, *117*, 102440. [[CrossRef](#)]
38. Wu, K.; Wang, Y.; Ye, Y.; Zhang, H.; Huang, G. Relationship between the built environment and the location choice of high-tech firms: Evidence from the Pearl River Delta. *Sustainability* **2019**, *11*, 3689. [[CrossRef](#)]
39. Onstein, A.T.; Ektesaby, M.; Rezaei, J.; Tavasszy, L.A.; van Damme, D.A. Importance of factors driving firms’ decisions on spatial distribution structures. *Int. J. Logist. Res. Appl.* **2020**, *23*, 24–43. [[CrossRef](#)]
40. Tao, Z.; Shuliang, Z. Collaborative innovation relationship in Yangtze River Delta of China: Subjects collaboration and spatial correlation. *Technol. Soc.* **2022**, *69*, 101974. [[CrossRef](#)]
41. Xu, B.; Yu, H.; Li, L. The impact of entrepreneurship on regional economic growth: A perspective of spatial heterogeneity. *Entrep. Reg. Dev.* **2021**, *33*, 309–331. [[CrossRef](#)]
42. Karahasan, B.C. Do new firms boost local innovation? Evidence from Turkey. *Int. Reg. Sci. Rev.* **2024**, *47*, 509–560. [[CrossRef](#)]

43. Hernández, R.C.; Camerin, F. The application of ecosystem assessments in land use planning: A case study for supporting decisions toward ecosystem protection. *Futures* **2024**, *161*, 103399. [CrossRef]
44. Wang, F.; Wong, W.-K.; Wang, Z.; Albasher, G.; Alsultan, N.; Fatemah, A. Emerging pathways to sustainable economic development: An interdisciplinary exploration of resource efficiency, technological innovation, and ecosystem resilience in resource-rich regions. *Resour. Policy* **2023**, *85*, 103747. [CrossRef]
45. Longato, D.; Cortinovis, C.; Balzan, M.; Geneletti, D. A method to prioritize and allocate nature-based solutions in urban areas based on ecosystem service demand. *Landsc. Urban Plan.* **2023**, *235*, 104743. [CrossRef]
46. Goodchild, M.F. What problem? Spatial autocorrelation and geographic information science. *Geogr. Anal.* **2009**, *41*, 411–417. [CrossRef]
47. Getis, A.; Ord, J.K. The analysis of spatial association by use of distance statistics. *Geogr. Anal.* **1992**, *24*, 189–206. [CrossRef]
48. Ord, J.K.; Getis, A. Local spatial autocorrelation statistics: Distributional issues and an application. *Geogr. Anal.* **1995**, *27*, 286–306. [CrossRef]
49. Ripley, B.D. Modelling spatial patterns. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **1977**, *39*, 172–192. [CrossRef]
50. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Routledge: London, UK, 2018.
51. Hohl, A.; Zheng, M.; Tang, W.; Delmelle, E.; Casas, I. Spatiotemporal point pattern analysis using Ripley’s K function. In *Geospatial Data Science Techniques and Applications*; Taylor & Francis Group: Oxford, UK, 2017; pp. 155–176.
52. Maiti, A.; Zhang, Q.; Sannigrahi, S.; Pramanik, S.; Chakraborti, S.; Cerda, A.; Pilla, F. Exploring spatiotemporal effects of the driving factors on COVID-19 incidences in the contiguous United States. *Sustain. Cities Soc.* **2021**, *68*, 102784. [CrossRef]
53. Yuan, J.; Wang, X.; Feng, Z.; Zhang, Y.; Yu, M. Spatiotemporal Variations of Aerosol Optical Depth and the Spatial Heterogeneity Relationship of Potential Factors Based on the Multi-Scale Geographically Weighted Regression Model in Chinese National-Level Urban Agglomerations. *Remote Sens.* **2023**, *15*, 4613. [CrossRef]
54. Priya Varshini, A.; Anitha Kumari, K.; Varadarajan, V. Estimating Software Development Efforts Using a Random Forest-Based Stacked Ensemble Approach. *Electronics* **2021**, *10*, 1195. [CrossRef]
55. Epifanio, I. Intervention in prediction measure: A new approach to assessing variable importance for random forests. *BMC Bioinform.* **2017**, *18*, 1–16. [CrossRef]
56. Louppe, G.; Wehenkel, L.; Sutura, A.; Geurts, P. Understanding variable importances in forests of randomized trees. In *Advances in Neural Information Processing Systems*; The MIT Press: Cambridge, MA, USA, 2013; Volume 26.
57. Fotheringham, A.S.; Yang, W.; Kang, W. Multiscale geographically weighted regression (MGWR). *Ann. Am. Assoc. Geogr.* **2017**, *107*, 1247–1265. [CrossRef]
58. Fotheringham, A.S.; Brunson, C.; Charlton, M. Geographically weighted regression. *Sage Handb. Spat. Anal.* **2009**, *1*, 243–254.
59. Dong, Y.; Guo, B.; He, D.; Liao, X.; Zhang, Z.; Wu, X. Industrial transformation and urban economic efficiency evolution: An empirical study of the Yangtze River economic belt. *Int. J. Environ. Res. Public Health* **2022**, *19*, 4154. [CrossRef] [PubMed]
60. Qin, C.; Huo, N.; Chong, Z. Intercity rail transit and integrated development of the pearl river delta urban cluster: Based on the perspective of network analysis. *Chin. J. Urban Environ. Stud.* **2015**, *3*, 1550024. [CrossRef]
61. Wei, H.; Nian, M.; Li, L. China’s strategies and policies for regional development during the period of the 14th five-year plan. *Chin. J. Urban Environ. Stud.* **2020**, *8*, 2050008. [CrossRef]
62. Chang, K.; Zhang, H.; Li, B. The impact of digital economy and industrial agglomeration on the changes of industrial structure in the Yangtze River Delta. *J. Knowl. Econ.* **2023**, 1–21. [CrossRef]
63. Bryson, J.R.; Daniels, P.W. *The Handbook of Service Industries*; Edward Elgar Publishing: Camberley, UK, 2007.
64. Díez-Vial, I.; Fernández-Olmos, M. Knowledge spillovers in science and technology parks: How can firms benefit most? *J. Technol. Transf.* **2015**, *40*, 70–84. [CrossRef]
65. Hailu, T.; Chebo, A.K.K. The Role of Industrial Parks Entrepreneurial Ecosystem in Strengthening Ventures’ Capability: Evidence from Ethiopian Small Manufacturing Enterprises. 2021. Available online: <https://www.researchsquare.com/article/rs-1049854/v1> (accessed on 25 April 2024).
66. Tonelli, M.; Dalglish, C. The role of transport infrastructure in facilitating the survival and growth of micro-enterprises in developing economies. In Proceedings of the 2012 Australian Centre for Entrepreneurship Research and DIANA Conference (ACERE DIANA), Fremantle, Australia, 3–5 February 2012; p. 111.
67. Wang, H.-Q.; Liang, L.-Q. How Do Housing Prices Affect Residents’ Health? New Evidence from China. *Front. Public Health* **2022**, *9*, 816372. [CrossRef] [PubMed]
68. Xu, H.; Zhang, D.; Zhang, Z. The impact of COVID-19 on international trade in China-industry review in the YRD and the PRD. *J. Educ. Humanit. Soc. Sci.* **2023**, *8*, 1763–1769. [CrossRef]
69. Li, X.; Tan, Y.; Xue, D. From world factory to global city-region: The dynamics of manufacturing in the Pearl River Delta and its spatial pattern in the 21st century. *Land* **2022**, *11*, 625. [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.