

Article

# An Integrated Duranton and Overman Index and Local Duranton and Overman Index Framework for Industrial Spatial Agglomeration Pattern Analysis

Yupu Huang <sup>1</sup> , Li Zhuo <sup>1,2,\*</sup>  and Jingjing Cao <sup>3</sup>

<sup>1</sup> Guangdong Provincial Engineering Research Center for Public Security and Disaster, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510006, China; huangyp66@mail2.sysu.edu.cn

<sup>2</sup> Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China

<sup>3</sup> College of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou 510665, China; caojj@gpnu.edu.cn

\* Correspondence: zhuoli@mail.sysu.edu.cn

**Abstract:** Accurately measuring industrial spatial agglomeration patterns is crucial for promoting regional economic development. However, few studies have considered both agglomeration degrees and cluster locations of industries. Moreover, the traditional multi-scale cluster location mining (MCLM) method still has limitations in terms of accuracy, parameter setting, calculation efficiency, etc. This study proposes a new framework for analyzing industrial spatial agglomeration patterns, which uses the Duranton and Overman (DO) index for estimating agglomeration degrees and a newly developed local DO (LDO) index for mining cluster locations. The MCLM-LDO method was proposed by incorporating the LDO index into the MCLM method, and it was validated via comparisons with three baseline methods based on two synthetic datasets. The results proved that the MCLM-LDO method can achieve accuracies of 0.945 and 1 with computational times of 0.15 s and 0.11 s on two datasets, which are superior to existing MCLM methods. The proposed framework was further applied to analyze the spatial agglomeration patterns of the industry of computer, communication, and other electronic equipment manufacturing in Guangdong Province, China. The results showed that the framework gives a more holistic perspective of spatial agglomeration patterns, which can serve as more meaningful references for industrial sustainable development.

**Keywords:** industrial spatial agglomeration; multi-scale cluster location mining (MCLM); Duranton and Overman (DO) index; distance-based method; micro-data; Guangdong Province



**Citation:** Huang, Y.; Zhuo, L.; Cao, J. An Integrated Duranton and Overman Index and Local Duranton and Overman Index Framework for Industrial Spatial Agglomeration Pattern Analysis. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 116. <https://doi.org/10.3390/ijgi13040116>

Academic Editors: Jiangfeng She, Jun Zhu, Min Yang and Wolfgang Kainz

Received: 8 February 2024

Revised: 17 March 2024

Accepted: 26 March 2024

Published: 29 March 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

Industrial spatial agglomeration refers to the geospatial concentration of interconnected firms [1], representing a worldwide phenomenon driven by economies of scale [2–4]. A suitable industrial agglomeration pattern, which considers both the agglomeration degree and the cluster location, can facilitate firms in benefiting from the industrial agglomeration effect and acquiring clear advantages in terms of cost, productivity, and innovation environment [5–10]. It serves as a crucial force for the high-quality and sustainable development of regions and cities [11–15]. A comprehensive analysis of industrial agglomeration patterns helps evaluate the performance of existing policies and formulate new ones, thereby promoting the development of a more suitable industrial agglomeration pattern [10,16–18]. Therefore, a comprehensive analysis of industrial agglomeration patterns is necessary. Previous studies have suggested that rapidly and accurately measuring agglomeration degrees and cluster locations is a prerequisite for comprehensive analysis of industrial spatial agglomeration patterns [19–21]. Nonetheless, industrial agglomeration degrees and cluster locations are interrelated and have significant spatial heterogeneity across different

regions, periods, and industry types, rendering the rapid and accurate measurement of both challenging [22–26].

Scholars in the early days primarily focused on measuring agglomeration degrees of industries at the administrative unit scale [17,22]. Specifically, an industrial agglomeration degree is typically quantified by constructing an index that reflects the deviation or proportion of the industry against all industries based on statistical data in administrative units [17]. However, traditional indexes are unable to accurately reflect agglomeration degrees, such as locational entropy, the Thiel index, spatial Gini coefficient, Herfindahl index, and EG index [13,15,27–31]. This is mainly because these indexes were primarily designed for a fixed spatial scale [17], which makes them inevitably influenced by the zoning scheme of the administrative unit, i.e., the existence of the Modifiable Areal Unit Problem (MAUP) [22,32,33].

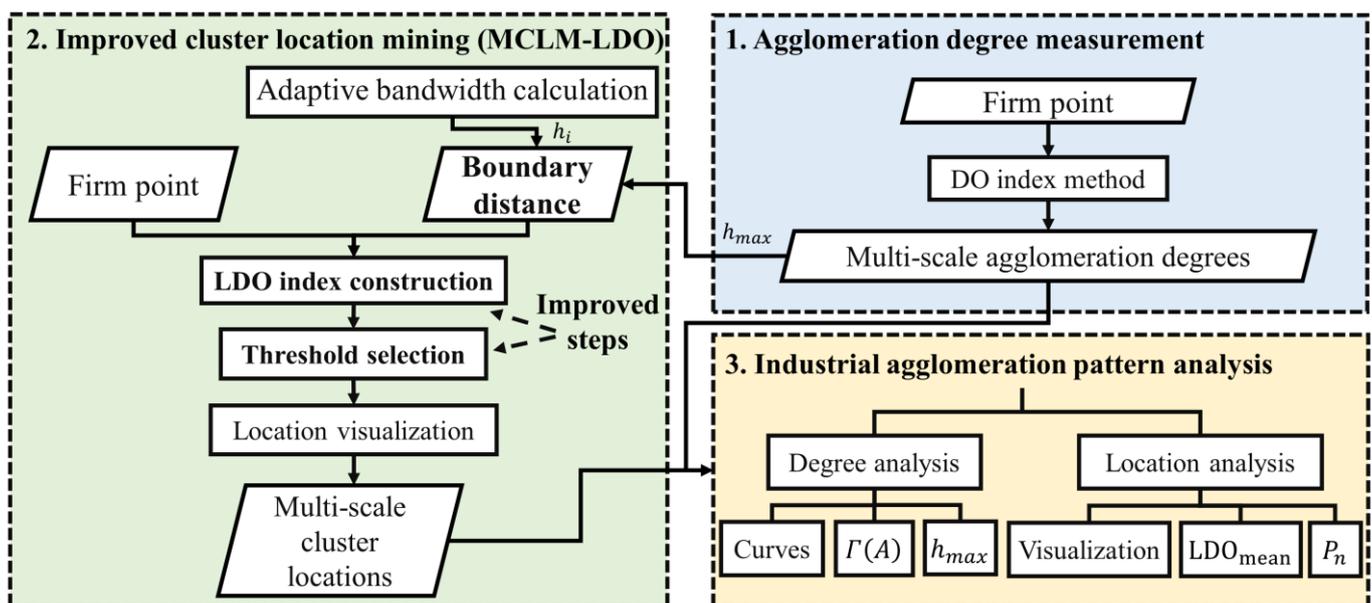
Subsequent studies proposed distance-based methods for measuring the agglomeration degrees of industries [22]. These methods can overcome the limitation of the fixed spatial scale and the MAUP [34,35] and produce more accurate agglomeration degrees [17,22]. Specifically, the distance-based method includes two steps, namely, calculating an average index representing the proximity of point-pair distances of firms to distance thresholds and comparing it with confidence intervals to obtain a curve representing the multi-scale agglomeration degree. Commonly used indexes of the distance-based agglomeration degree measuring method include Ripley's K-function (i.e., the K-function) [36] and its variant, L-function [37,38], as well as the Duranton and Overman (DO) index [20]. Among these indexes, the K-function and L-function, which count the numbers of point-pair distances of firms that are less than different distance thresholds to measure proximity, tend to overestimate the spatial scale of the agglomeration distribution. This is because the result of large spatial scales contains small spatial scales, which consequently results in a cumulative effect [20,22]. The DO index employs a kernel function to represent the proximity and address the cumulative effect in the K-function and L-function [20,39]. Although several indexes, such as the M-function [40,41] and W-function [42], have emerged, they cannot entirely substitute the DO index because of their low computational efficiency, rendering it challenging to apply them to massive datasets [22]. Thereby, the distance-based method based on the DO index has become the predominant method for measuring industrial agglomeration degrees of regions [17]. It has been widely used in studies of industrial agglomeration in the country, urban agglomeration, and inner city levels [19,20,23,39,43–51]. Nevertheless, the DO index method still faces challenges, such as a lack of explanations on the result curves, low accessibility of micro-data on firms, and ignorance of cluster locations [21,22,52].

Several methods have been adopted to obtain accurate industrial cluster locations, such as the spatial scan statistic [21,53] and the kernel density estimation [54–56]. However, these methods cannot simultaneously obtain cluster locations and their relationships at multiple scales, because these cluster locations may vary at different scales [24]. To further address these limitations, Buzard et al. proposed a Multi-scale Cluster Location Mining (MCLM) method based on the Local K-function (LK), i.e., MCLM-LK. This method has provided promising results for mining cluster locations in research and development labs and breweries in America [35,57]. Specifically, the MCLM-LK method counts the number of industry firms within a given boundary distance parameter for each firm and then adopts a local test to identify core firms for constituting industrial cluster locations. Although the MCLM-LK method combined with the K-function method offers an effective way of simultaneously measuring agglomeration degrees and mining cluster locations of industries [35,57,58], it still has several limitations. First, the LK exhibits a significant cumulative effect, which may lead to identical calculation results for neighboring firms and add biases to cluster locations. Second, the boundary distance parameter in the MCLM-LK method depends on subjective experience inputs, rendering it difficult to jointly analyze agglomeration degrees and cluster locations. Third, the computational cost is very high because the local test requires an additional 999 calculations for each firm at each spatial scale.

This study aims to propose a new framework for analyzing multi-scale industrial spatial agglomeration patterns that simultaneously consider agglomeration degrees and cluster locations. In the proposed framework, agglomeration degrees are calculated by using the DO index, while cluster locations are estimated based on a newly developed local DO (LDO) index. By incorporating the LDO index, the traditional MCLM method is improved to provide more accurate cluster location mining results with higher efficiency and a more objective parameter setting. The proposed MCLM-LDO method will be compared with three baseline methods based on two synthetic datasets. The integrated DO index and LDO index (DO-LDO) framework will be applied to analyze the spatial agglomeration patterns of the industry of computer, communication, and other electronic equipment manufacturing in Guangdong Province of China from 2000 to 2022. The remainder of this study is organized as follows. Section 2 describes the proposed framework and Section 3 analyzes the experimental results, followed by the discussion and conclusions in Sections 4 and 5.

## 2. Methodology

As shown in Figure 1, the DO-LDO framework of this study mainly comprises three parts. First, the multi-scale agglomeration degrees are measured by using the DO index method. Then, to obtain multi-scale cluster locations, a novel LDO index is constructed, and an MCLM-LDO method is proposed by introducing the LDO index into the MCLM. Thirdly, a comprehensive analysis of industrial spatial agglomeration patterns is performed from the dual perspectives of agglomeration degrees and cluster locations.



**Figure 1.** The overall structure of the proposed integrated Duranton and Overman (DO) index and Local Duranton and Overman (LDO) index (DO-LDO) framework. Note that MCLM-LDO refers to the Multi-scale Cluster Location Mining method based on LDO index,  $h_{max}$  and  $h_i$  represent boundary distance parameters for global and local spatial scales, respectively,  $\Gamma(A)$  denotes the sum over all distances of the agglomeration degree,  $LDO_{mean}$  denotes the average density of firms in the cluster, and  $P_n$  denotes the percentage of the number of firms in the cluster to the total number of firms in the industry.

### 2.1. Multi-Scale Agglomeration Degree Measurement Based on DO Index Method

This study adopted the widely used DO index method to measure agglomeration degrees because it often provides a satisfactory result for the multi-scale agglomeration degrees of industries [17]. The DO index of the investigated industry  $A$  at the distance threshold  $d$  can be calculated by Equation (1), denoted as  $DO^{obs}(A, d)$ . Generally, the

sequence of distance thresholds starts at 0 and increments by one or one-tenth of the distance unit. The maximum distance threshold will be set at the quartile (e.g., median or lower quartile) of point-pair distances of firms or the area diameter [20,45].

$$DO(A, d) = \frac{1}{hN_A(N_A - 1)} \sum_{i=1}^{N_A-1} \sum_{j=i+1}^{N_A} e^{-\left(\frac{d-d_{i,j}}{\sqrt{2}h}\right)^2} \quad (1)$$

where  $N_A$  is the number of firms in industry  $A$ ,  $d_{i,j}$  is the Euclidean distance between firm  $i$  and firm  $j$  (unless otherwise specified, distance in this study refers to Euclidean distance), and  $h$  is the optimal bandwidth calculated using Silverman's method (Equation (2)) [59].

$$h = 0.9N_A^{-\frac{1}{5}} \min(d^{std}, \frac{d^{Q1}}{1.34}) \quad (2)$$

where  $d^{std}$  and  $d^{Q1}$  represent the standard deviation and lower quartile of point-pair distances of all firms in the industry  $A$ .

An upper global confidence band for the random distribution of industry  $A$  is then generated through a counterfactual test [19,20]. The main steps include the following:

(1) A Monte Carlo sample approach is used to create  $m$  simulations, with  $N_A$  firms in each simulation ( $m$  generally ranges from 100 to 1000) from all background industries (e.g., manufacturing);

(2) DO indexes of these simulations,  $DO_m^{sim}(A, d)$ , are calculated;

(3) For each  $d$ , all  $DO_m^{sim}(A, d)$  of simulations are ranked in descending order, and the initial upper global confidence band is selected as the 5-th percentile;

(4) It is determined whether the number of simulations greater than the upper global confidence band for at least one  $d$  exceeds 5% $m$ ;

(5) If so, a larger value, i.e., the  $(5 - \frac{1}{m})$ -th percentile, is taken as the new upper global confidence band to perform step (4), otherwise, a determined upper global confidence band  $\overline{DO^{sim}(A, d)}$  is achieved.

For example, there are 1000 simulations created and sorted, which can be expressed as  $DO_1^{sim}(A, d) > DO_2^{sim}(A, d) > \dots > DO_{1000}^{sim}(A, d)$  at each  $d$ . Among them,  $DO_{50}^{sim}(A, d)$  denotes the initial upper global confidence band. If step (4) is not achieved, then use  $DO_{49}^{sim}(A, d)$  as a new upper global confidence band and continue with step (4) until it is achieved.

Finally, the localization index  $\Gamma(A, d)$  at different spatial scales of industry  $A$  is calculated by Equation (3), and a localization index curve can be created for quantifying the multi-scale agglomeration degree [20,43].

$$\Gamma(A, d) = \max\left(DO^{obs}(A, d) - \overline{\overline{DO^{sim}(A, d)}}, 0\right) \quad (3)$$

where  $\Gamma(A, d)$  is greater than 0 at  $d$ , the spatial scale at which industry  $A$  is agglomerated, the larger  $\Gamma(A, d)$  indicates the higher agglomeration degree and the stronger industrial agglomeration effect.

## 2.2. LDO Index Construction for MCLM-LDO Method

To address the limitations of the MCLM-LK method, this study proposes an MCLM-LDO method by introducing a novel LDO index into the MCLM. The MCLM-LDO method first constructs the LDO index and proposes an objective determination approach for its distance parameter, enhancing an accurate measurement of the agglomeration degrees of individual firms. Then, the threshold selection is used to identify firms with higher LDO indexes as core firms, i.e., individual firms with significantly high agglomeration degrees are identified, dramatically reducing the computation time compared to the local test, and multi-scale cluster locations are subsequently visualized based on the core firms. Finally, an evaluation in terms of both the accuracy and efficiency of the proposed MCLM-LDO method was performed.

### 2.2.1. Construction of LDO Index

The preliminary step of the MCLM is to measure the agglomeration degrees of individual firms by using a suitable index. A traditional index, LK, was commonly used by counting the number of firms within a specified boundary distance parameter  $\hat{h}$  (Equation (4)). However, it results in the shortcoming that neighboring firms have identical degrees.

$$\text{LK}_i(A, \hat{h}) = C_i(\hat{h}) \quad (4)$$

where  $C_i(\hat{h})$  denotes the number of firms in industry  $A$  whose distance from firm  $i$  is not greater than  $\hat{h}$ .

The DO index has addressed the cumulative effect of the K-function, thereby enabling a more accurate measurement of industrial agglomeration degrees in a region [20,39]. This study, therefore, constructs a local version of the DO index to address the shortcomings of the LK function. Specifically, this study adopts a kernel function instead of the counting method in the LK, and it constructs a novel index (the LDO index) to obtain neighboring firms with different degrees for better measuring the agglomeration degrees of individual firms. For a firm  $i$  in industry  $A$ , its LDO index when the peak and boundary distance parameters are  $\hat{d}$  and  $\hat{h}$ , respectively, is denoted as  $\text{LDO}_i(A, \hat{d}, \hat{h})$  and can be described as Equation (5).

$$\text{LDO}_i(A, \hat{d}, \hat{h}) = \frac{1}{\hat{h}N_A} \sum_{j \neq i}^{N_A} e^{-\left(\frac{\hat{d}-d_{ij}}{\sqrt{2}\hat{h}}\right)^2} \quad (5)$$

The peak distance represents the distance where the agglomeration degree is maximized, and the boundary distance represents the farthest distance where the agglomeration degree is detectable. Given that agglomeration degrees of individual firms decay with distance [6,24], the peak distance  $\hat{d}$  in this study is set to 0, and the LDO index can be further simplified as Equation (6). Inputting different  $\hat{h}$  values will obtain the LDO index of a firm at different spatial scales.

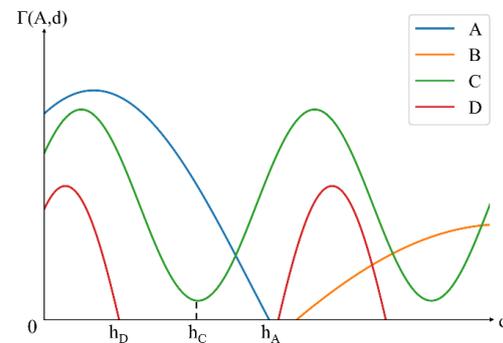
$$\text{LDO}_i(A, \hat{h}) = \frac{1}{\hat{h}N_A} \sum_{j \neq i}^{N_A} e^{-\left(\frac{d_{ij}}{\hat{h}}\right)^2} \quad (6)$$

### 2.2.2. Determination of Boundary Distance Parameters

The boundary distance parameter  $\hat{h}$  of the LDO index needs to be determined for performing Equation (6). Currently, the boundary distance parameter of LK relies on subjective experience for determination, ignoring the heterogeneity of the boundary distance parameter in different regions, periods, and industry types. This makes it difficult to accurately mine multi-scale cluster locations and their relationships. In this study, both global and local spatial scales are considered to improve the accuracy of mining multi-scale cluster locations and their relationships. Combined with the advantage that the localization index curve is a superior characterization for multi-scale agglomeration degrees, an approach is proposed to determine the boundary distance parameter at a global scale (denoted as  $h_{max}$ ), and global scopes of industrial agglomeration can be depicted. For the local location of small cluster, an adaptive approach is used to determine the boundary distance parameter at a local scale (denoted as  $h_i$ ).

The global scale parameter of the agglomerated industry,  $h_{max}$ , is determined as the smaller distance between the first decline of the curve to 0 and the first curve trough. For example, in Figure 2, the localization index curves of industries A, C, and D exhibit  $\Gamma(A, d)$  greater than 0 within short distances, indicating agglomeration. The corresponding  $h_A$ ,  $h_C$ , and  $h_D$  represent the  $h_{max}$  for these industries. In contrast, the localization index curve of industry B is greater than 0 at long distances, indicating dispersion [20], with no  $h_{max}$  determined. The local scale parameter  $h_i$  refers to Siverman's optimal bandwidth of the DO index method and is calculated using Equation (2). The difference here is that  $d^{std}$  and  $d^{Q1}$

represent the standard deviation and lower quartile of distances of firm  $i$  from the other firms in industry  $A$ . Through calculation, each firm  $i$  will have its  $h_i$ , which can detect clusters at the local scale different from the global scale.



**Figure 2.** Examples of localization index curves of industries A, B, C, and D.

### 2.2.3. Core Firm Identification Based on Threshold Selection

Based on the determination of two objective boundary distances, LDO indexes of all firms were calculated (Equation (6)), and cluster locations can be obtained by identifying the core firms, i.e., the firms with higher LDO indexes. The MCLM-LK method is generally limited by computational inefficiency due to the local test, that is, calculating 999 simulated values for each firm through the Monte Carlo simulation and identifying the firms whose observed value is higher than all of their simulated values to be the core firms [35]. In this study, a threshold selection approach is proposed to identify core firms at multiple scales. The LDO index of a firm complying with Equation (7) is considered as a core firm.

$$\text{LDO}_i(A, \hat{h}) \geq \max(\mu, 2\sigma) \quad (7)$$

where  $\mu$  and  $\sigma$  denote the mean and standard deviation of LDO indexes of firms in industry  $A$  with input  $\hat{h}$ . When the distances between core firms are greater than  $\hat{h}$ , they are considered to belong to different clusters. If the number of a cluster is greater than or equal to  $0.5\%N_A$  [54], the cluster is selected as the finally estimated cluster.

### 2.2.4. Cluster Location Visualization

This study uses different visualization approaches for the global and local scales to compare cluster locations. For the global spatial scale, the global agglomeration boundary is constructed using the minimum bounding rectangle of core firms, where the firm with the largest LDO index is the agglomeration center of the industry. For the local spatial scale, the spatial distribution of core firms is directly used as the visualization results.

### 2.2.5. Performance Evaluation of MCLM-LDO Method

Due to the difficulty of obtaining real cluster locations to calculate accuracy indicators, two synthetic datasets characterizing common industrial spatial distribution patterns are used in this study. The aim is to evaluate the effectiveness of the MCLM-LDO method in identifying cluster locations and compare it with three baseline methods. The evaluation metrics include the accuracy and computational efficiency of methods.

The evaluation scheme is shown in Figure 3. The MCLM-LDO method with three baseline methods is applied to two synthetic datasets separately to obtain estimated firm types at the global scale (i.e., inputting  $h_{max}$  from the localization index curve) and uses real firm types as the reference to evaluate both in terms of efficiency and accuracy. The efficiency indicator uses the computation time, and the accuracy adopts three indicators from the confusion matrix. *Recall* denotes the rate of the number of core firms (i.e., the firms in clusters) correctly estimated to the number of real core firms (Equation (8)); *Specificity* denotes the rate of the number of sparse firms (i.e., the firms beyond clusters) correctly

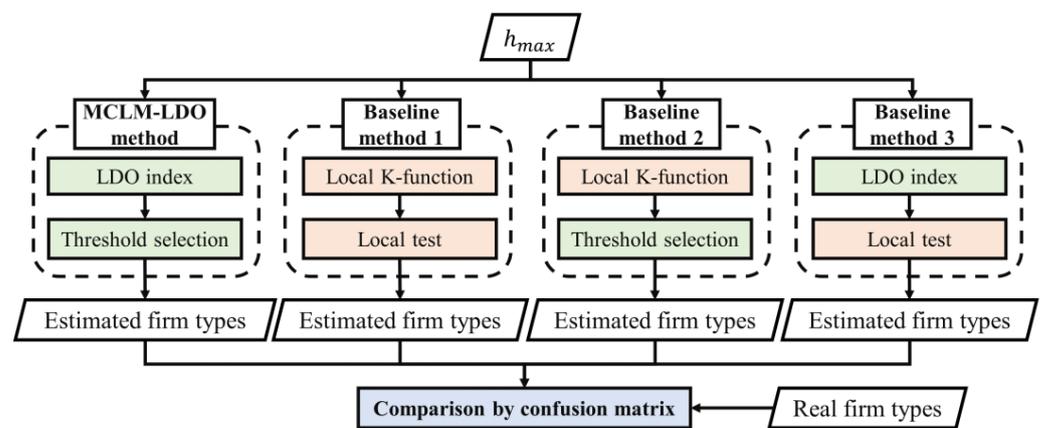
estimated to the number of real sparse firms (Equation (9)); and *Accuracy* denotes the rate of the number of firms correctly estimated to the total number of firms (Equation (10)). The larger the three indicators, the higher the accuracy of the method.

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$Specificity = \frac{TN}{FP + TN} \quad (9)$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (10)$$

where *TP* is the number of core firms correctly estimated by the method, *TN* is the number of sparse firms correctly estimated by the method, and *FP* as well as *FN* are the number of firms incorrectly estimated by the method, respectively.



**Figure 3.** Diagram of the evaluation scheme for the MCLM-LDO method. Note that green rectangles represent improved steps while orange rectangles represent steps of the MCLM-LK method.

Baseline method 1 is the MCLM-LK method proposed by Buzard et al. [35]. The MCLM-LK method initially calculates observed values (i.e., agglomeration degrees of individual firms) of the industry based on LK (Equation (4)), and then the core firms are identified by the local test. To evaluate the effectiveness of each improved step of the MCLM-LDO method, baseline method 2 and baseline method 3 were constructed using the step in the MCLM-LDO method replacing the step in the MCLM-LK method.

### 2.3. Industrial Spatial Agglomeration Pattern Analysis from Dual Perspectives

This study finally employed the proposed framework to analyze the industrial spatial agglomeration patterns of Guangdong Province from 2000 to 2022, using micro-data on firms in the computer, communication, and other electronic equipment manufacturing industry. The multi-scale agglomeration degree of the industry is analyzed using the localization index curve and its derivations, including the aggregated agglomeration degree  $\Gamma(A)$  (Equation (11)) and the maximum spatial scale of agglomeration  $h_{max}$ .

$$\Gamma(A) = \sum_d \Gamma(A, d) \quad (11)$$

where  $\Gamma(A)$  greater than 0 indicates that the spatial distribution of industry *A* is agglomerated in the region; otherwise, it is considered random or dispersed. A larger  $\Gamma(A)$  signifies a more significant agglomeration of industry *A* in the region.

Subsequently, industrial cluster locations are analyzed at the global and local scales. Additionally, the trends of cluster are quantified by calculating the average LDO index (denoted as  $LDO_{mean}$ , representing the average firm density of each cluster) and the percentage of firms within each cluster relative to  $N_A$  (denoted as  $P_n$ ).

### 3. Experiments and Analysis

#### 3.1. Datasets

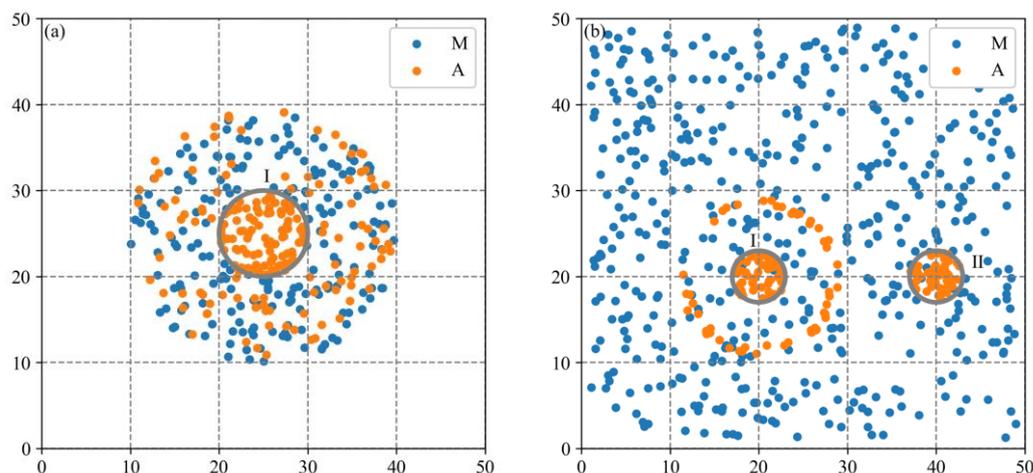
##### 3.1.1. Synthetic Datasets

This study constructs the synthetic datasets by considering two industries. As listed in Table 1, A represents the investigated industry, M represents other background industries, and firms were randomly generated on a  $50 \times 50$  area.

**Table 1.** Descriptions of industry A in the synthetic datasets.

Datasets	Number of Clusters	Scale of Clusters	Number of Core Firms	Number of Sparse Firms
Dataset 1	1	10	100	100
Dataset 2	2	6	100	50

Synthetic Dataset 1 depicts an industrial spatial distribution of a single cluster surrounded by sparse firms (Figure 4a). Cluster I, with a scale (i.e., circular diameter) of 10, hosts 100 core firms of industry A. A ring, spanning from 5 to 15 away from the center of Cluster I, accommodates 100 sparse firms of industry A and 200 firms of industry M.



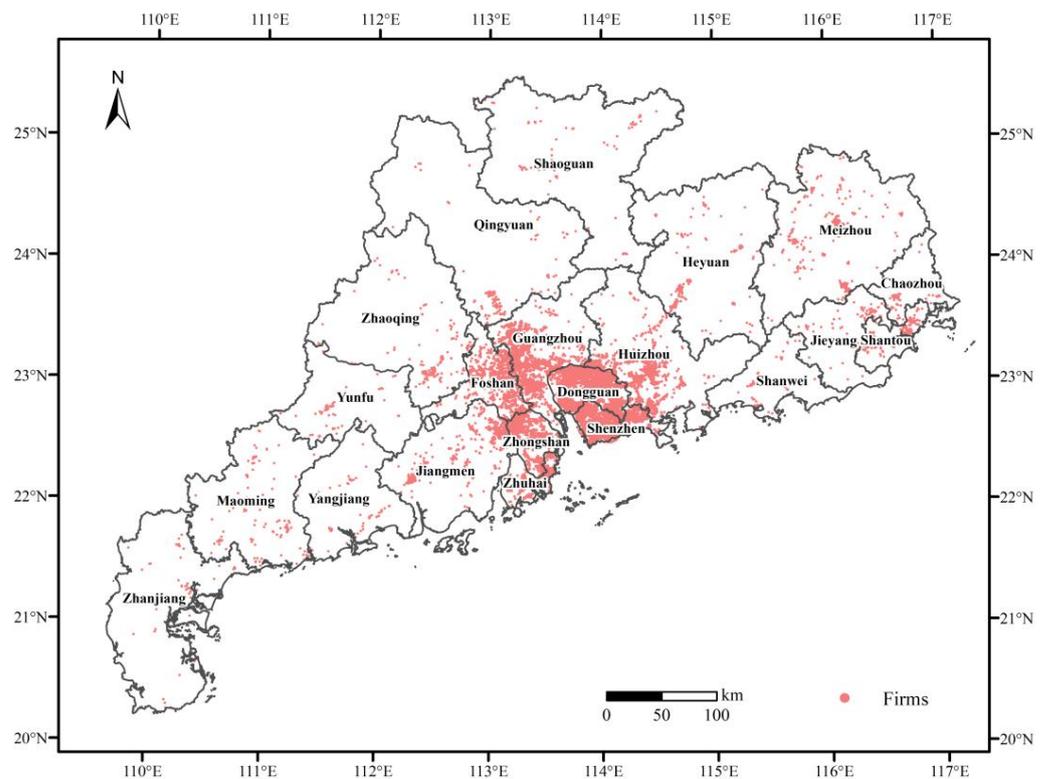
**Figure 4.** The industrial spatial distributions of multiple clusters in (a) Synthetic Dataset 1 and (b) Synthetic Dataset 2.

Synthetic Dataset 2 depicts an industrial spatial distribution of multiple clusters, with sparse firms at a distance from a cluster (Figure 4b). Both Cluster I and Cluster II have a scale of 6, each hosting 50 core firms of industry A. A ring, located 8 to 9 away from the center of Cluster I, accommodates 50 sparse firms of industry A. Moreover, industry M comprises 500 firms randomly distributed across the entire area.

##### 3.1.2. Actual Dataset

This study adopts an actual dataset of firms in Guangdong Province, China, encompassing 21 cities, obtained from Qichacha, a Chinese enterprise information query platform ("<https://www.qcc.com/>" (accessed on 11 April 2023)). This dataset represents the point data of surviving firms in the computer, communication, and other electronic equipment manufacturing industry in Guangdong Province from 1949 to 2022 (Figure 5). The latitude and longitude of the firms were obtained using the geocoder service of Baidu Maps. A total of 75,275 firms were ultimately used after three preprocessing steps using Python and ArcGIS 10.4 software, including eliminating invalid data, standardizing attributes, and correcting city-level locations (i.e., manually coding firms whose cities from the geocoder service differed from the "city" field). According to the "National Economic Industry Classification Standard (GB/T 4754-2017) [60]" of China, the industry is coded as C39. Since a

significant portion of firms in this industry were established after 2000, the analysis of this study mainly concentrates on six specific years, i.e., 2000, 2005, 2010, 2015, 2020, and 2022.



**Figure 5.** Spatial distribution of firms in the C39 industry in Guangdong Province, China, in 2022. Note that the Dongsha Islands of China are not shown considering that there are no firms.

Guangdong Province is the most industrially and economically developed province in China. For the past 35 years, Guangdong Province has always been the top province in China in terms of Gross Domestic Product (GDP). The central region in Guangdong is the Pearl River Delta, comprising nine cities: Guangzhou, Foshan, Zhaoqing, Shenzhen, Dongguan, Huizhou, Zhuhai, Zhongshan, and Jiangmen. Presently, Guangdong Province grapples with a serious industrial and economic imbalance between its central region and the other regions [61,62]. Preventing the excessive concentration of industries in the central region and fostering a coordinated regional development strategy are pressing concerns [61]. In addition, in 2022, the value added by the C39 industry in Guangdong Province significantly surpassed that of other industries. Therefore, an appropriate framework to analyze the agglomeration patterns of this industry is crucial for the sustainable development of this region.

### 3.2. Performance Analysis of MCLM-LDO Method on Synthetic Data

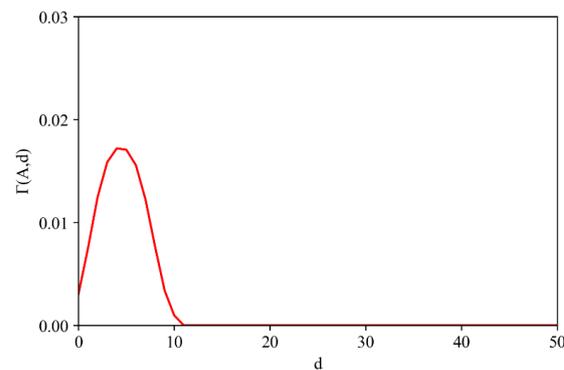
To obtain the  $h_{max}$  to input into the MCLM-LDO method and three baseline methods, the localization index curves for two synthetic datasets were calculated by inputting a sequence of distance thresholds  $D = [0, 1, \dots, 25]$  and  $m = 1000$  to the DO index method.

#### 3.2.1. Spatial Agglomeration Analysis on Synthetic Dataset 1

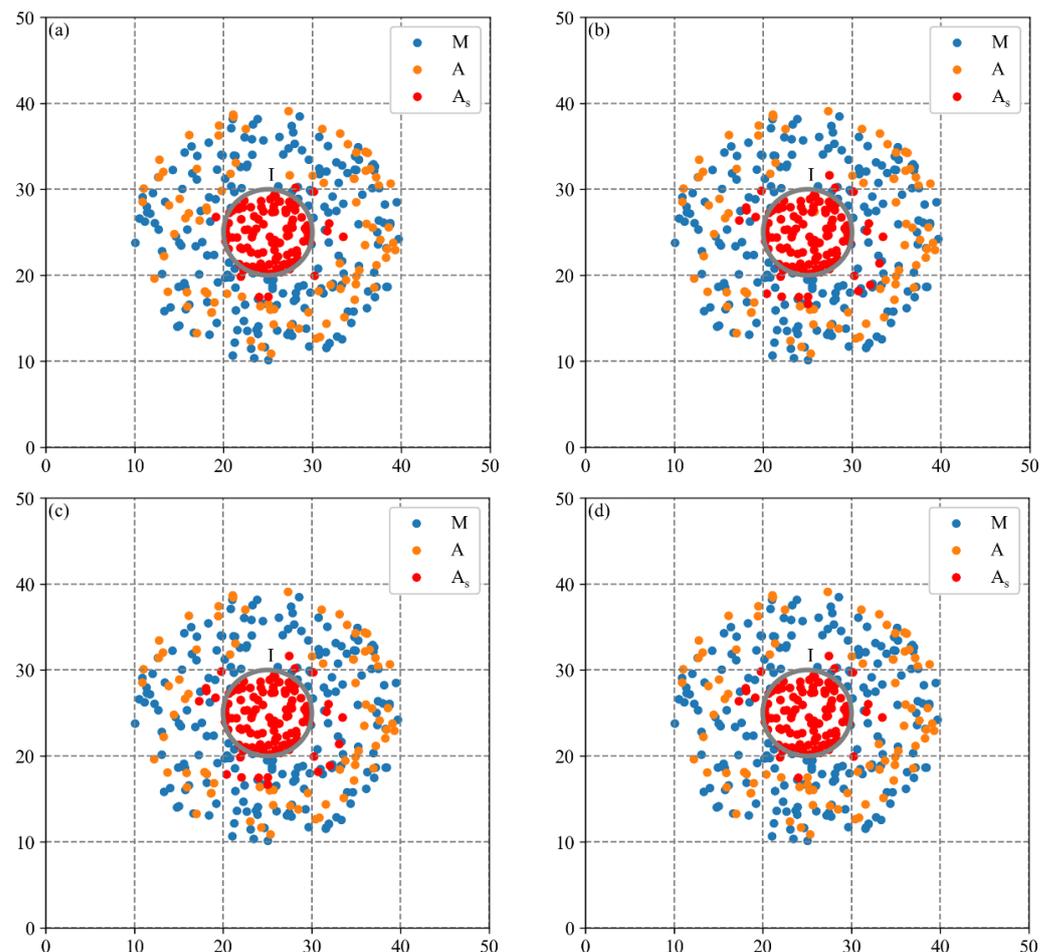
As shown in Figure 6, the  $h_{max}$  of industry A is 11, closely matching the scale of Cluster I (10), demonstrating the accuracy and objectivity of  $h_{max}$  as the boundary distance parameter.

Figure 7 illustrates the distributions of estimated firm types for the MCLM-LDO method and three baseline methods obtained by inputting  $\hat{h} = 11$ . The core firms in industry A are denoted as  $A_s$ , while the remainder are sparse firms. Some sparse firms around Cluster I are misidentified as core firms by all methods. As depicted in Table 2,

the MCLM-LDO method correctly identified 100 core firms and 89 sparse firms, and its  $Recall = 1$ ,  $Specificity = 0.89$ , and  $Accuracy = 0.945$ , which outperforms the other three methods in all indicators, achieving a computation time of only 0.15 s. The  $Recall$  of each method is similar, indicating that these methods have comparable abilities to correctly identify core firms. The  $Specificity$  of the two methods using the LDO index (i.e., MCLM-LDO method and baseline method 1) reaches 0.89 and 0.85, respectively, significantly surpassing the other methods using LK (i.e., baseline method 1 and baseline method 2). This suggests that the LDO index is more effective in correctly identifying sparse firms.



**Figure 6.** The localization index  $\Gamma(A, d)$  of Synthetic Dataset 1.



**Figure 7.** Spatial distribution of estimated firm types on Synthetic Dataset 1 by inputting  $\hat{h} = 11$  to (a) MCLM-LDO method, (b) baseline method 1, (c) baseline method 2, and (d) baseline method 3.

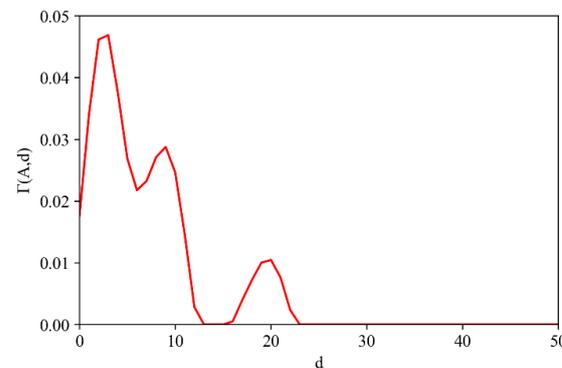
**Table 2.** Comparison of accuracy and computational efficiency of different methods based on Synthetic Dataset 1.

Indicators	Baseline Method 1	Baseline Method 2	Baseline Method 3	MCLM-LDO Method
Recall	0.99	1	1	1
Specificity	0.78	0.78	0.85	0.89
Accuracy	0.885	0.89	0.925	0.945
Computational time (s)	24.21	0.16	24.82	0.15

In summary,  $h_{max}$  proves to be a suitable boundary distance parameter, and the methods using the LDO index, especially the MCLM-LDO method, more accurately pinpoint the cluster location compared to using the LK.

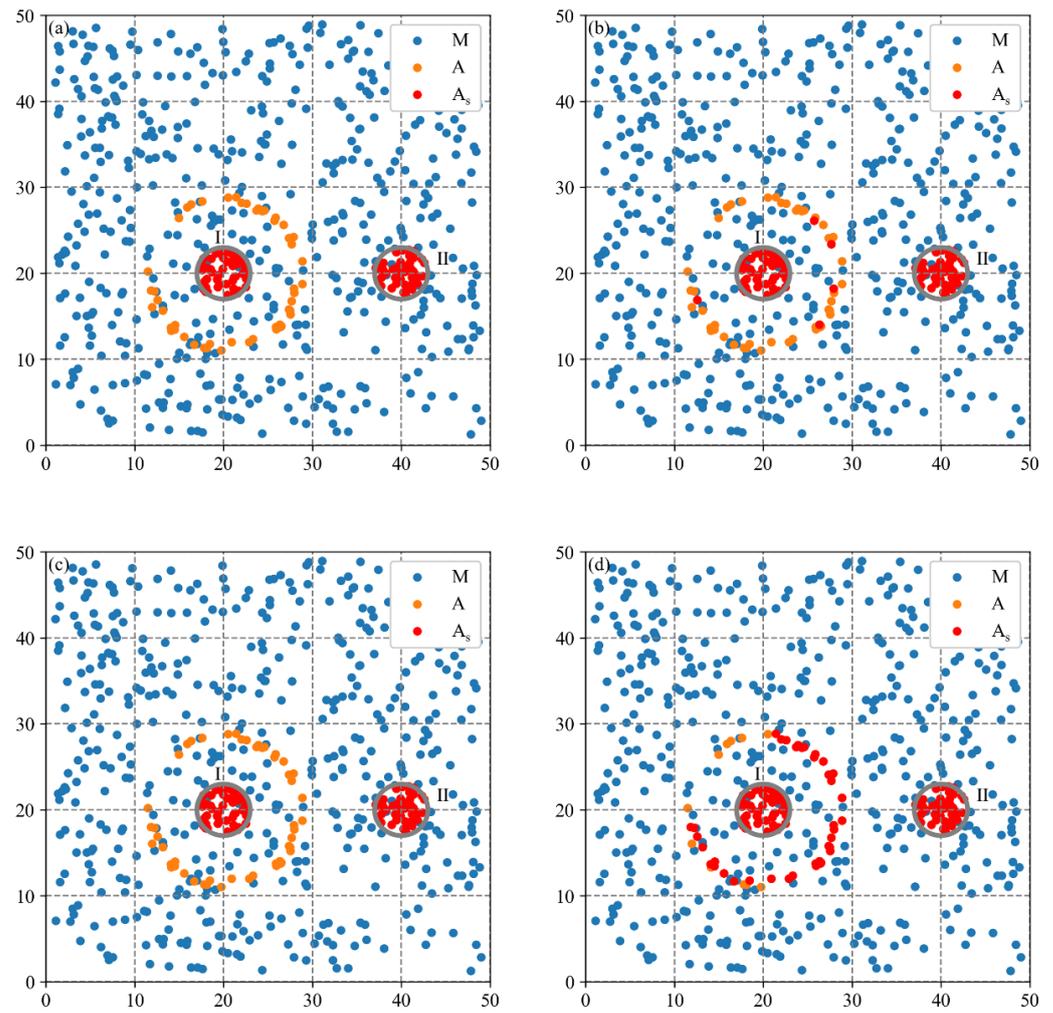
### 3.2.2. Spatial Agglomeration Analysis on Synthetic Dataset 2

Figure 8 displays the localization index curve of Synthetic Dataset 2, revealing three curve crests. The  $h_{max}$ , located 6 from the first curve trough, is consistent with the scales of the two clusters on Synthetic Dataset 2.

**Figure 8.** The localization index  $\Gamma(A, d)$  of Synthetic Dataset 2.

By inputting  $\hat{h} = 6$  into the MCLM-LDO method and three baseline methods, the distributions of estimated firm types are shown in Figure 9, with the corresponding indicators presented in Table 3. The MCLM-LDO method and baseline method 2 correctly identified core firms and sparse firms by using the threshold selection, as indicated by their *Accuracy* reaching 1. In contrast, the other two baseline methods, both using the local test, have a *Specificity* that is less than 1, indicating that they misidentify some sparse firms as core firms. This is because the local test aims to identify areas where industry A is more intensive than industry M, i.e., the surrounding area of the ring on this synthetic dataset. Additionally, compared to baseline method 1, the lower *Specificity* of baseline method 3 is attributed to the LDO index focusing more on the distribution of neighboring firms, while the LK considers all firms within a distance of 6 equally. In the surrounding area of the ring, industry A is more intensive relative to industry M within a distance of 1, but it has a similar density as industry M within a distance of 6.

Furthermore, the computation time for the threshold selection (0.11 s for MCLM-LDO and 0.12 s for baseline method 2) is significantly lower than that of the local test (18.6 s and 19.1 s for baseline method 1 and baseline method 3, respectively). Overall, the threshold selection outperforms the local test in both accuracy and computational efficiency.



**Figure 9.** Spatial distribution of estimated firm types on Synthetic Dataset 2 by inputting  $\hat{h} = 6$  to (a) MCLM-LDO method, (b) baseline method 1, (c) baseline method 2, and (d) baseline method 3.

**Table 3.** Comparison of accuracy and computational efficiency of different methods based on Synthetic Dataset 2.

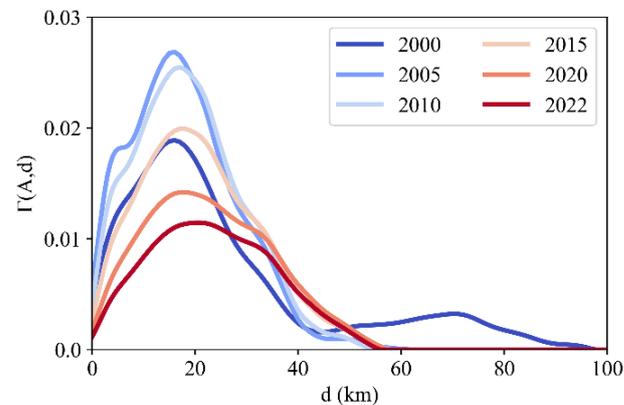
Indicators	Baseline Method 1	Baseline Method 2	Baseline Method 3	MCLM-LDO Method
Recall	1	1	1	1
Specificity	0.9	1	0.22	1
Accuracy	0.967	1	0.74	1
Computational time (s)	18.6	0.12	19.1	0.11

### 3.3. Application of Integrated DO Index and LDO Index Framework on Actual Dataset

#### 3.3.1. Industrial Agglomeration Degree Analysis of Guangdong Province

By inputting a sequence of distance thresholds  $D = [0 \text{ km}, 1 \text{ km}, \dots, 100 \text{ km}]$  and  $m = 400$  to the DO index method, Figure 10 and Table 4 illustrate the evolution of agglomeration degrees for the C39 industry in Guangdong Province from 2000 to 2022. The  $\Gamma(A)$  of the C39 industry in Guangdong Province is consistently greater than 0 and the  $h_{max}$  exceeds 45 km in all years, indicating that the industrial agglomeration at the large spatial scale persists from 2000 to 2022. By collectively analyzing  $\Gamma(A, d)$ ,  $\Gamma(A)$ , and  $h_{max}$ , the evolution of the spatial agglomeration patterns of the C39 industry in Guangdong Province can be preliminarily divided into three periods. The first period is from 2000 to 2005, where  $\Gamma(A)$  rises sharply from 0.58 to 0.68, with a moderate change in  $h_{max}$ , reflecting

rapid agglomeration of the industry and a stable spatial scale of agglomeration. The second period is from 2005 to 2015, where  $\Gamma(A)$  declines gradually from 0.68 to 0.58, while  $h_{max}$  expands rapidly from 46 km to 57 km, signifying the start of industry dispersal and a noticeable growth in the spatial scale of agglomeration. The third period is from 2015 to 2022, where  $\Gamma(A)$  decreases dramatically from 0.58 to 0.38, with a small change in  $h_{max}$ , indicating a rapid dispersion of the industry but a stable spatial scale of agglomeration.



**Figure 10.** The evolution of the localization index  $\Gamma(A, d)$  of the C39 industry in Guangdong Province from 2000 to 2022.

**Table 4.** The evolution of the aggregated agglomeration degree  $\Gamma(A)$  and the maximum spatial scale of agglomeration  $h_{max}$  of the C39 industry in Guangdong Province from 2000 to 2022.

Indicators	2000	2005	2010	2015	2020	2022
$\Gamma(A)$	0.58	0.68	0.67	0.58	0.47	0.38
$h_{max}$ (km)	45	46	54	57	57	55

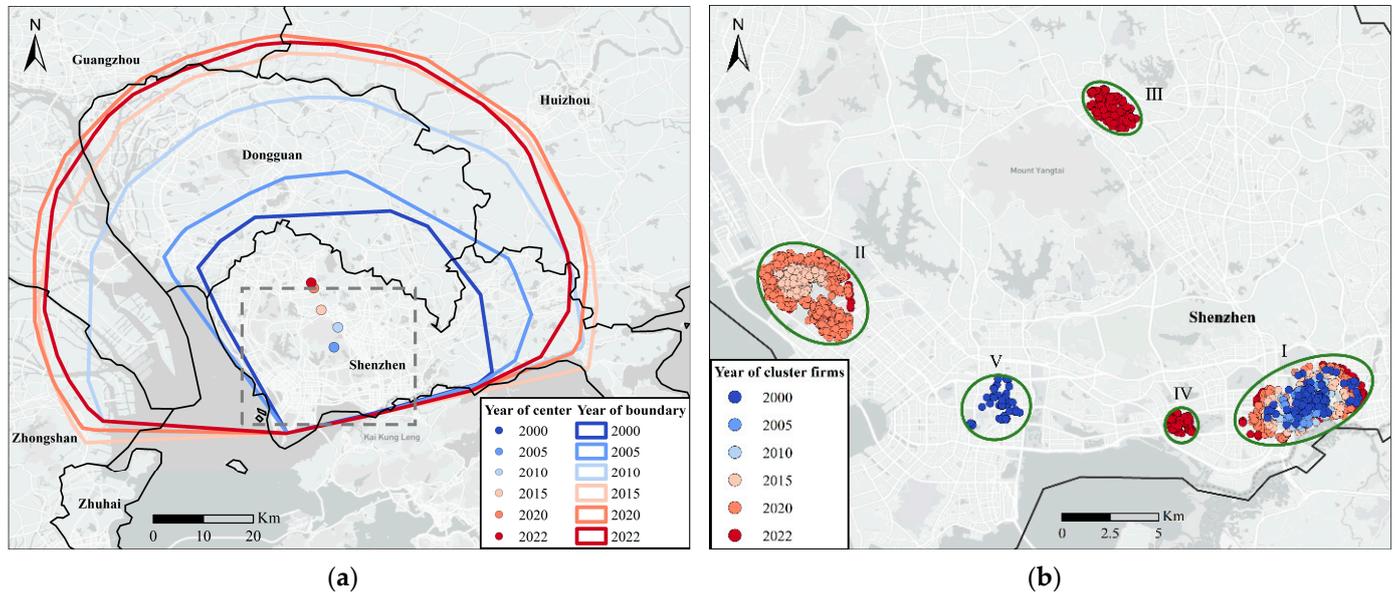
### 3.3.2. Industrial Cluster Location Analysis of Guangdong Province

Figure 11 illustrates the evolution of cluster locations for the C39 industry in Guangdong Province from 2000 to 2022 obtained by using the MCLM-LDO method. On the global scale, the agglomeration boundary of the industry experienced a slight expansion in 2005. Subsequently, the boundary rapidly expanded outward from 2005 to 2015, centered around Shenzhen City, crossing the Pearl River in 2010. This formed a spatial agglomeration pattern of urban agglomerations with integrated development on the east and west coasts. After 2015, the boundary tended to stabilize, and the agglomeration center gradually moved closer to Dongguan City along transportation arteries. This implies a change in the industrial spatial agglomeration patterns inside the boundary, warranting further analysis.

As shown in Figure 11b and Table 5, at the local scale, the local industrial clusters in all years are primarily distributed in the coastal area of Shenzhen City and all are located inside the global agglomeration boundary. In 2005, Cluster V was insignificant, while the  $LDO_{mean}$  of Cluster I rapidly rose to 0.0435, indicating a significant increase in the firm density of Cluster I. Between 2015 and 2020, Cluster II developed strongly and its area expanded greatly, with  $P_n$  growing from 1.08% to 2.91%. In 2022, Clusters III and IV emerged, and Cluster I continued to have the highest  $LDO_{mean}$  and  $P_n$  among the clusters. These results reflect changes in the industrial spatial patterns inside the global agglomeration boundary.

Overall, the three periods of the agglomeration patterns of the C39 industry in Guangdong Province from 2000 to 2022 can be summarized as follows. (1) In the first period of a stabilizing agglomeration pattern from 2000 to 2005, the industry rapidly clustered in Shenzhen City with a stable boundary, forming a significant pattern of single Cluster I (i.e., the cluster of the Huaqiang North Road) in the local area. (2) In the second period of an expanding dispersion pattern from 2005 to 2015, the industry tended to disperse, and its global agglomeration boundary expanded considerably, generating an agglomeration

pattern of urban agglomerations. (3) In the third period of an internal dispersion pattern from 2015 to 2022, the industry dispersion accelerated, manifesting in the movement of the agglomeration center to Dongguan City and the emergence of multiple significant local clusters.



**Figure 11.** The evolution of cluster locations of the C39 industry in Guangdong Province from 2000 to 2022: (a) global agglomeration boundary and center and (b) local cluster (inside the gray dotted line in (a)).

**Table 5.** The evolution of local industrial clusters of the C39 industry in Guangdong Province from 2000 to 2022. Note that “-” means that the cluster is not significantly clustered in that year.

Cluster Numbers	LDO <sub>mean</sub>						P <sub>n</sub> (%)					
	2000	2005	2010	2015	2020	2022	2000	2005	2010	2015	2020	2022
I	0.0287	0.0435	0.0347	0.0210	0.0124	0.0108	7.45	6.97	5.90	4.24	3.26	3.26
II	-	-	-	0.0096	0.0068	0.0062	-	-	-	1.08	2.91	2.64
III	-	-	-	-	-	0.0049	-	-	-	-	-	0.60
IV	-	-	-	-	-	0.0049	-	-	-	-	-	0.58
V	0.0203	-	-	-	-	-	3.66	-	-	-	-	-

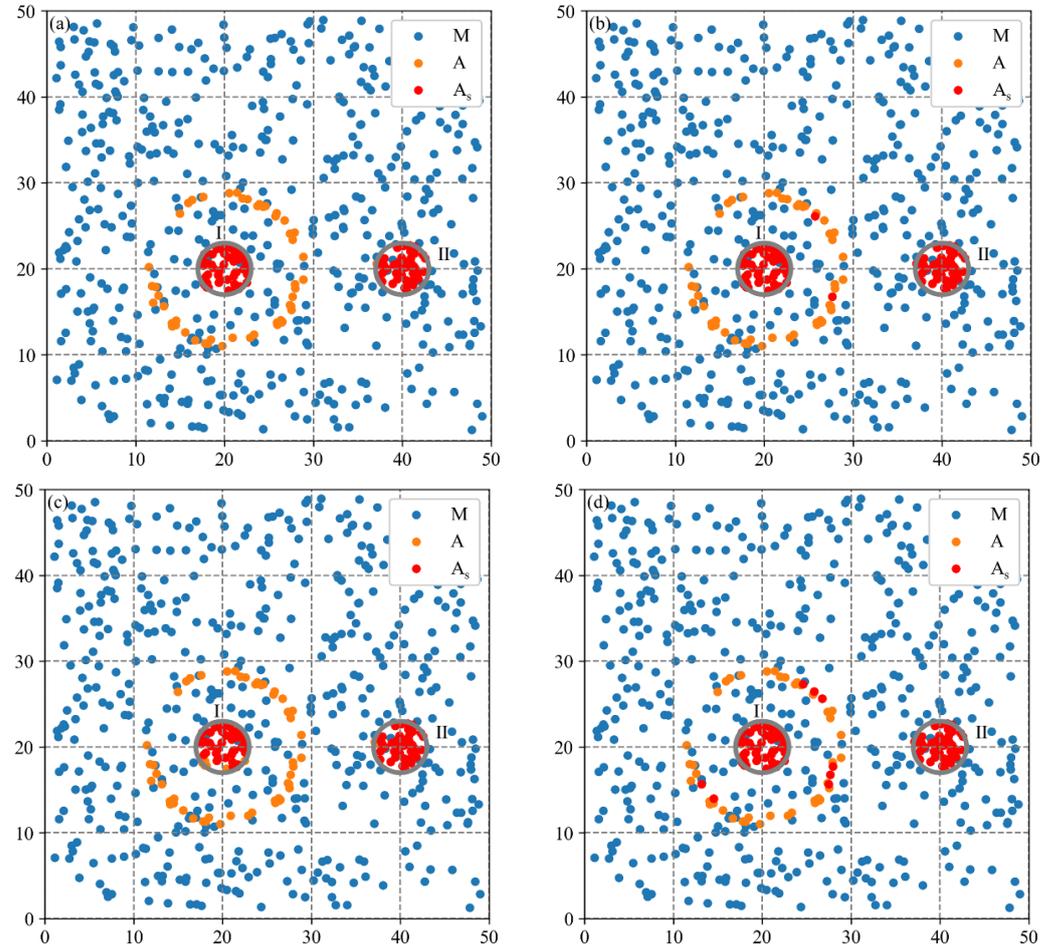
As shown in Table 5, since 2015, among these clusters, the most significant clusters of the C39 industry in Guangdong Province are Cluster I and Cluster II (i.e., the cluster of the Airport Economic Zone of Shenzhen). Cluster I remains consistently significant, but its P<sub>n</sub> has been decreasing, indicating that the increase in the number of firms inside the cluster has been lower than that outside the cluster. This implies that Cluster I is unable to accommodate more new firms. Therefore, this cluster needs to rely on the transformation and upgrading of existing firms in the future to continuously develop. In addition, Cluster II is noteworthy due to its rapid development.

#### 4. Discussion

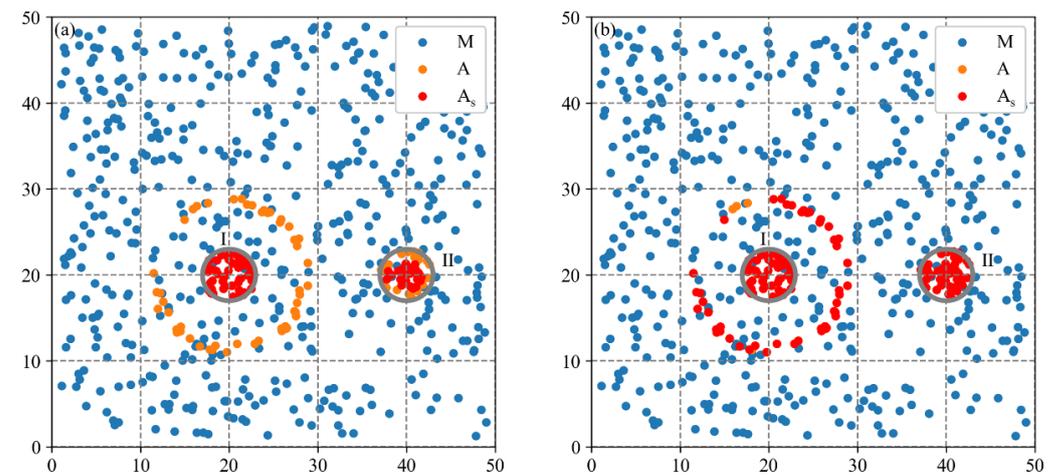
##### 4.1. Sensitivity Analysis of Distance Parameters

The sensitivity of distance parameters of the MCLM-LDO method, i.e., the boundary distance  $\hat{h}$  and the peak distance  $\hat{d}$ , is analyzed in this section. Taking Synthetic Dataset 2 as an example, the additional parameters,  $\hat{h} = 4$  and  $\hat{h} = 8$ , are input into the MCLM-LDO method and three baseline methods, and the distributions of the estimated firms are shown in Figures 12 and 13, respectively. Overall, the results identified by the four methods vary

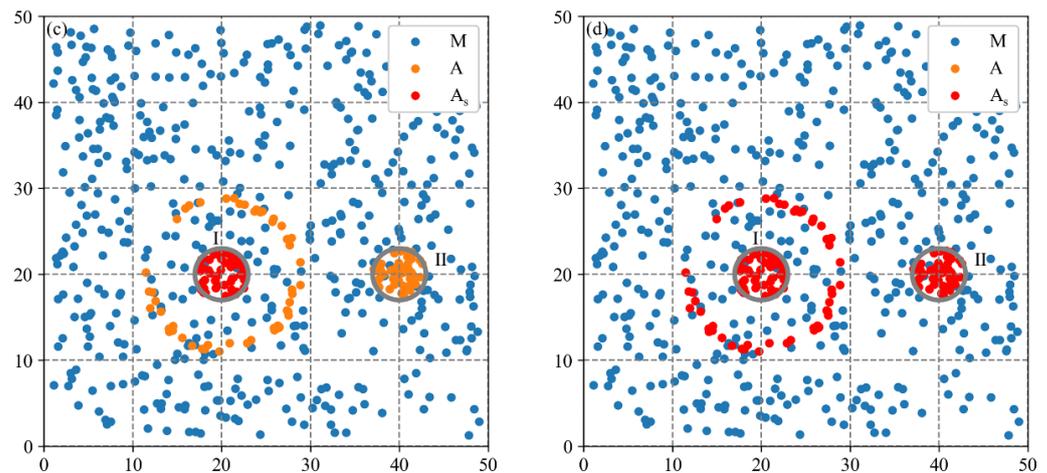
when  $\hat{h}$  is different. As shown in Table 6, the three indicators of the MCLM-LDO method varied insignificantly and remained considerably higher than the other three methods, demonstrating that the MCLM-LDO method has superior robustness and can achieve the best result when  $\hat{h} = h_{max}$ .



**Figure 12.** Spatial distribution of estimated firm types on Synthetic Dataset 2 by inputting  $\hat{h} = 4$  to the (a) MCLM-LDO method, (b) baseline method 1, (c) baseline method 2, and (d) baseline method 3.



**Figure 13.** Cont.



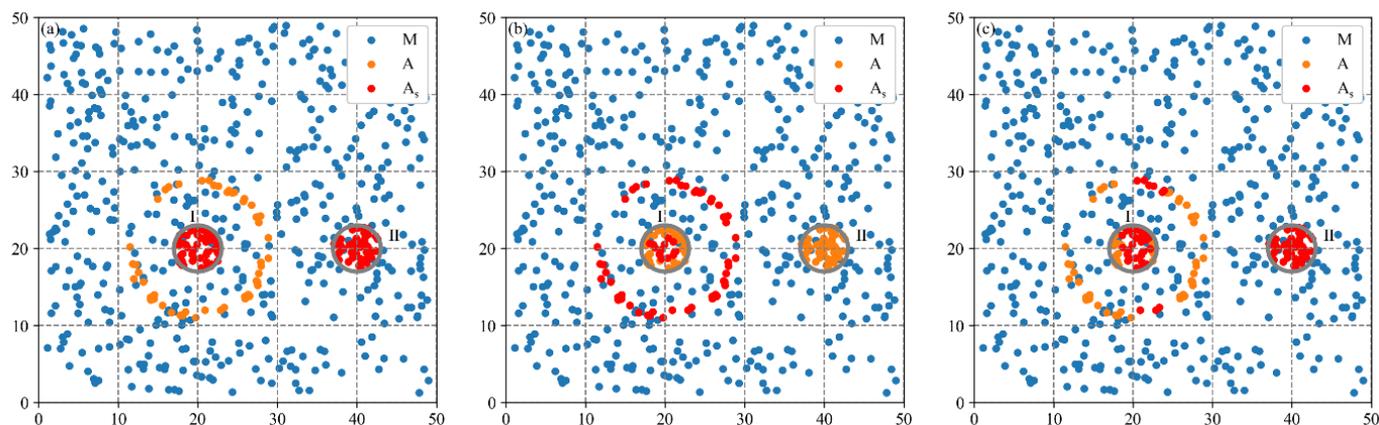
**Figure 13.** Spatial distribution of estimated firm types on Synthetic Dataset 2 by inputting  $\hat{h} = 8$  to the (a) MCLM-LDO method, (b) baseline method 1, (c) baseline method 2, and (d) baseline method 3.

**Table 6.** Sensitivity analysis of the boundary distance parameter on Synthetic Dataset 2.

Indicators	Input $\hat{h}=4$			Input $\hat{h}=8$				
	Baseline Method 1	Baseline Method 2	Baseline Method 3	MCLM-LDO Method	Baseline Method 1	Baseline Method 2	Baseline Method 3	MCLM-LDO Method
Recall	1	0.89	1	0.99	1	0.5	1	0.77
Specificity	0.96	1	0.84	1	0.08	1	0	1
Accuracy	0.987	0.927	0.947	0.993	0.693	0.67	0.667	0.85

For the peak distance  $\hat{d}$ , this study set it to 0 in the MCLM-LDO method. When  $\hat{d}$  is greater than 0, the MCLM-LDO method mines points with larger impacts that are spaced a certain distance. For example, some plants, when spaced at a certain distance from each other, will develop better [6], and the MCLM-LDO method can be used to identify plants that match this condition. Furthermore, previous studies have argued the conclusion that multiple crests of the localization index curve imply the existence of multiple clusters in the industry [43,45]. By using the MCLM-LDO method with  $\hat{d} > 0$ , the correctness of this conclusion of previous studies can be discussed in this section.

The localization index curve of Synthetic Dataset 2 has multiple crests (Figure 8).  $\hat{d}$  is set to be the distance of the maximum value of the three crests, and  $\hat{h}$  is set to be 2. The combinations of distance parameters  $(\hat{d}, \hat{h}) = (3, 2), (9, 2),$  and  $(20, 2)$  are inputted into the MCLM-LDO method, and their results are shown in Figure 14. In this case, the core firms of each combination explain why the localization index curve crests at those distances. (1) The first crest at a short distance identifies core firms located inside Cluster I and Cluster II, representing the cluster locations of industry A. (2) The core firms identified by the second crest are not located inside clusters of industry A but are exactly near 9 from Cluster I; therefore, they have higher LDO indexes. (3) The third crest identifies core firms mostly inside two clusters of industry A, which are close to each other by 20. Generally, the crests of the localization index curve may be generated both by the distance between multiple clusters and by sparse firms at some distance from the cluster. Thus, this study argues that the localization index curve having multiple crests does not necessarily mean that there are multiple clusters.



**Figure 14.** Spatial distribution of estimated firm types by using the MCLM-LDO method on Synthetic Dataset 2 by inputting (a) the first curve crest ( $\hat{d} = 3$ ,  $\hat{h} = 2$ ), (b) second curve crest ( $\hat{d} = 9$ ,  $\hat{h} = 2$ ), and (c) third curve crest ( $\hat{d} = 20$ ,  $\hat{h} = 2$ ).

#### 4.2. Improvement of the MCLM-LDO Method

The advantages of the proposed MCLM-LDO method lie in adopting a new approach to determine boundary distance parameters, constructing the LDO index, and proposing the threshold selection approach.

(1) The determination of boundary distance parameters is data-driven and objective, facilitating generalization across various regions, periods, and industry types, which is challenging for previous methods reliant on a priori knowledge for parameter settings [35].

(2) The LDO index can effectively differentiate the agglomeration degrees of neighboring firms compared to the traditional LK. Consequently, the MCLM-LDO method mines cluster locations with greater accuracy than the MCLM-LK method.

(3) The threshold selection approach involves two steps, i.e., calculating the standard deviation and mean of the data and conducting a comparison. These steps have significantly improved the computational efficiency and applicability to the large-scale data. In contrast, the local test requires an additional 999 calculations for each firm at each spatial scale, resulting in highly inefficient computational processes. Furthermore, the local test bears the risk of misidentifying firms in sparse regions, leading to lower accuracy of the result.

Moreover, the MCLM-LK method used by Buzar et al. further constructs buffers, centered on core firms with a radius, as the final cluster [35,57], while this step is not included in the MCLM-LDO method. This is because the radius of buffers directly uses the boundary distance parameter [35], and in some cases, the results may be unreasonable. For example, if this step is used to generate clusters of Synthetic Dataset 2 (Figure 9), all sparse firms around Cluster I will be covered, resulting in a considerable overestimation of the results; for an industry with a large boundary distance parameter (e.g., C39 industry at the global scale in this study), the scope of the cluster derived from constructing buffers will be too large and thus meaningless.

#### 4.3. Applicability of the DO-LDO Framework

Most studies have analyzed industrial agglomeration patterns using the DO index method, obtaining accurate agglomeration degrees but paying insufficient attention to cluster location mining [17,22,23]. Although the MCLM-LK method has focused on cluster location mining, it still suffers from accuracy and efficiency shortcomings and is difficult to analyze with the agglomeration degree [57]. Based on the DO index method, this study proposed the MCLM-LDO method to improve the accuracy and efficiency of the MCLM method. Moreover, the proposed DO-LDO framework can effectively integrate the agglomeration degree and the cluster location to analyze the industrial agglomeration pattern comprehensively.

Based on the DO-LDO framework, this study presents findings regarding the evolution and the current status of the C39 industry (the computer, communication, and electronic equipment manufacturing industry) in Guangdong Province. By 2022, the industry has formed an integrated development along the east and west coasts, with startups gravitating towards emerging local clusters within the global agglomeration scope, while original local clusters have reached saturation points. Overall, the C39 industry in Guangdong Province exhibits a pattern of "global single-core agglomeration and local multi-point diffusion". Policymakers can evaluate the policy effectiveness and formulate sustainable industrial and urban development policies based on the results obtained from applying our framework.

(1) While the supportive policies have alleviated the decline of the Huaqiang North Road cluster, startups increasingly dispersed into the surrounding areas, maintaining a crowding effect. Therefore, the cluster's policies should focus on promoting the upgrading and relocation of internal firms and relieving the population and land pressure to safeguard sustainable urban development.

(2) The initial effectiveness of the coordinated regional development strategy of Guangdong Province is evident. The C39 industry can be further distributed in Dongguan City or along the west coast of the Pearl River based on the Guangdong–Hong Kong–Macao Bridge and the Shenzhen–Zhongshan Corridor. It will mitigate excessive concentration in Shenzhen City and foster polycentric and sustainable industrial development.

#### 4.4. Extensibility, Limitations, and Future Work

Referring to the DO index [20], the LDO index also has a form that considers the weights of firms (Equation (12)).

$$\text{LDO}_i(A, \hat{d}, \hat{h}) = \frac{1}{\hat{h} \sum_{j \neq i}^{N_A} w_i w_j} \sum_{j \neq i}^{N_A} w_i w_j e^{-\left(\frac{\hat{d}-d_{ij}}{\sqrt{2\hat{h}}}\right)^2} \quad (12)$$

where  $w_i$  and  $w_j$  are the weights of firms  $i$  and  $j$ , such as the number of employees. When the number of small- and medium-scale firms has a greater quantity in the industry, an additive weighting form can be considered [48] (Equation (13)).

$$\text{LDO}_i(A, \hat{d}, \hat{h}) = \frac{1}{\hat{h} \sum_{j \neq i}^{N_A} (w_i + w_j)} \sum_{j \neq i}^{N_A} (w_i + w_j) e^{-\left(\frac{\hat{d}-d_{ij}}{\sqrt{2\hat{h}}}\right)^2} \quad (13)$$

The distance metric of the LDO index can consider other linkages between firms, such as economic, knowledge, and vertical industrial linkage. Road distance is also a metric that can be considered. If alternative distance-based methods emerge, a step similar to the LDO index construction can also be applied to mine cluster locations. Since the DO-LDO framework and the MCLM-LDO method are conducted based on pure point progress, it is applicable across different economic systems and industries if the input data are well developed.

The DO-LDO framework also facilitates the balance of efficiency and equity in industrial distribution to support high-quality and sustainable development. For example, analyzing industrial agglomeration patterns in ecologically fragile areas can prevent the further development of heavily polluting industries.

Nevertheless, the DO-LDO framework still faces several limitations. For example, the distance parameters and threshold of the MCLM-LDO method need to be adjusted for different domains, thereby improving the rationality of spatial point pattern analysis. It is important to note that alternative clustering methods may be more suitable in scenarios like text recognition and map clustering, where multiple classifications for all points are necessary [35,63]. Moreover, more statistical indicators can be added to quantify the level of development of industrial agglomeration in a more multidimensional way.

## 5. Conclusions

This study constructed a novel LDO index and proposed the MCLM-LDO method for industrial cluster location mining to address the limitations of the existing MCLM-LK method in terms of accuracy, parameter setting, and calculation efficiency. The DO-LDO framework was performed to comprehensively analyze the industrial multi-scale spatial agglomeration patterns in Guangdong Province of China from 2000 to 2022 by considering the dual perspective of agglomeration degrees and cluster locations. The main conclusions of this study are as follows.

(1) The proposed MCLM-LDO method can provide industrial cluster locations at the global and local scales and deepen the understanding of the localization index curve.

(2) By inputting the objective distance parameter, the evaluation of two synthetic datasets demonstrated that the MCLM-LDO method yields superior results in accuracy and computational efficiency, compared with other baseline methods.

(3) The spatial agglomeration patterns of the C39 industry in Guangdong Province from 2000 to 2022 include three periods, a stabilizing agglomeration pattern from 2000 to 2005, an expanding dispersion pattern from 2005 to 2015, and an internal dispersion pattern from 2015 to 2022.

These findings can provide a scientific reference for the sustainable planning of the industry and analyze the impacts and mechanisms of industrial agglomeration.

**Author Contributions:** Conceptualization, Li Zhuo, Yupu Huang, and Jingjing Cao; data curation, Li Zhuo and Yupu Huang; methodology, Yupu Huang; formal analysis, Yupu Huang; writing—original draft preparation, Yupu Huang; writing—review and editing, Li Zhuo and Jingjing Cao; visualization, Yupu Huang. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China (grant No. 41971372 and 42201353), Guangdong Basic and Applied Basic Research Foundation (grant No. 2022B1515130001), and Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (grant No. 311022009).

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the authors on request.

**Acknowledgments:** We thank the anonymous reviewers and all of the editors that participated in the revision process.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Martin, R.; Sunley, P. Deconstructing Clusters: Chaotic Concept or Policy Panacea? *J. Econ. Geogr.* **2003**, *3*, 5–35. [[CrossRef](#)]
- Glaeser, E.L.; Kallal, H.D.; Scheinkman, J.A.; Shleifer, A. Growth in Cities. *J. Polit. Econ.* **1992**, *100*, 1126–1152. [[CrossRef](#)]
- Guo, D.; Jiang, K.; Xu, C.; Yang, X. Industrial Clustering, Income and Inequality in Rural China. *World Dev.* **2022**, *154*, 105878. [[CrossRef](#)]
- Porter, M.E. Competitive Advantage, Agglomeration Economies, and Regional Policy. *Int. Reg. Sci. Rev.* **1996**, *19*, 85–90. [[CrossRef](#)]
- Steijn, M.P.A.; Koster, H.R.A.; Van Oort, F.G. The Dynamics of Industry Agglomeration: Evidence from 44 Years of Coagglomeration Patterns. *J. Urban Econ.* **2022**, *130*, 103456. [[CrossRef](#)]
- Fang, L.; Drucker, J. How Spatially Concentrated Are Industrial Clusters?: A Meta-Analysis. *J. Plan. Lit.* **2021**, *36*, 526–542. [[CrossRef](#)]
- de Groot, H.L.F.; Poot, J.; Smit, M.J. Which Agglomeration Externalities Matter Most and Why? *J. Econ. Surv.* **2016**, *30*, 756–782. [[CrossRef](#)]
- Peng, C.; Elahi, E.; Fan, B.; Li, Z. Effect of High-Tech Manufacturing Co-Agglomeration and Producer Service Industry on Regional Innovation Efficiency. *Front. Environ. Sci.* **2022**, *10*, 942057. [[CrossRef](#)]
- Xu, D.; Yu, B.; Liang, L. High-Tech Industrial Agglomeration and Urban Innovation in China's Yangtze River Delta Urban Agglomeration: From the Perspective of Industrial Structure Optimization and Industrial Attributes. *Complexity* **2022**, *2022*, 2555182. [[CrossRef](#)]
- Guo, X.; Guo, K.; Zheng, H. Industrial Agglomeration and Enterprise Innovation Sustainability: Empirical Evidence from the Chinese A-Share Market. *Sustainability* **2023**, *15*, 11660. [[CrossRef](#)]
- Du, H.; Ji, X.; Chuai, X. Spatial Differentiation and Influencing Factors of Water Pollution-Intensive Industries in the Yellow River Basin, China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 497. [[CrossRef](#)] [[PubMed](#)]

12. Zheng, H.; He, Y. How Does Industrial Co-Agglomeration Affect High-Quality Economic Development? Evidence from Chengdu-Chongqing Economic Circle in China. *J. Clean Prod.* **2022**, *371*, 133485. [[CrossRef](#)]
13. Zhang, H.; Zhang, J.; Song, J. Analysis of the Threshold Effect of Agricultural Industrial Agglomeration and Industrial Structure Upgrading on Sustainable Agricultural Development in China. *J. Clean Prod.* **2022**, *341*, 130818. [[CrossRef](#)]
14. Zhang, L.; Mu, R.; Hu, S.; Zhang, Q.; Wang, S. Impacts of Manufacturing Specialized and Diversified Agglomeration on the Eco-Innovation Efficiency—A Nonlinear Test from Dynamic Perspective. *Sustainability* **2021**, *13*, 3809. [[CrossRef](#)]
15. Liu, X.; Zhang, X.; Sun, W. Does the Agglomeration of Urban Producer Services Promote Carbon Efficiency of Manufacturing Industry? *Land Use Pol.* **2022**, *120*, 106264. [[CrossRef](#)]
16. Chen, Z.; Xu, W.; Zhao, Z. The Assessment of Industrial Agglomeration in China Based on NPP-VIIRS Nighttime Light Imagery and POI Data. *Remote Sens.* **2024**, *16*, 417. [[CrossRef](#)]
17. Chain, C.P.; Santos, A.C.D.; Castro, L.G.D.; Prado, J.W.D. Bibliometric Analysis of The Quantitative Methods Applied to The Measurement of Industrial Clusters. *J. Econ. Surv.* **2019**, *33*, 60–84. [[CrossRef](#)]
18. Borana, S.L.; Yadav, S.K. Urban Land-Use Susceptibility and Sustainability—Case Study. In *Water, Land, and Forest Susceptibility and Sustainability*; Chatterjee, U., Pradhan, B., Kumar, S., Saha, S., Zakwan, M., Fath, B.D., Fiscus, D., Eds.; Academic Press: Cambridge, MA, USA, 2023; Volume 2, pp. 261–286.
19. Brakman, S.; Garretsen, H.; Zhao, Z. Spatial Concentration of Manufacturing Firms in China. *Pap. Reg. Sci.* **2017**, *96*, S179–S205. [[CrossRef](#)]
20. Duranton, G.; Overman, H.G. Testing for Localization Using Micro-Geographic Data. *Rev. Econ. Stud.* **2005**, *72*, 1077–1106. [[CrossRef](#)]
21. Zhang, X.; Yao, J.; Sila-Nowicka, K.; Song, C. Geographic Concentration of Industries in Jiangsu, China: A Spatial Point Pattern Analysis Using Micro-Geographic Data. *Ann. Reg. Sci.* **2021**, *66*, 439–461. [[CrossRef](#)]
22. Marcon, E.; Puech, F. A Typology of Distance-Based Measures of Spatial Concentration. *Reg. Sci. Urban Econ.* **2017**, *62*, 56–67. [[CrossRef](#)]
23. Huang, Y.; Sheng, K.; Sun, W. Influencing Factors of Manufacturing Agglomeration in the Beijing-Tianjin-Hebei Region Based on Enterprise Big Data. *Acta Geogr. Sin.* **2022**, *77*, 1953–1970. [[CrossRef](#)]
24. Rosenthal, S.S.; Strange, W.C. How Close Is Close? The Spatial Reach of Agglomeration Economies. *J. Econ. Perspect.* **2020**, *34*, 27–49. [[CrossRef](#)]
25. Verstraten, P.; Verweij, G.; Zwaneveld, P.J. Complexities in the Spatial Scope of Agglomeration Economies. *J. Reg. Sci.* **2019**, *59*, 29–55. [[CrossRef](#)]
26. Cainelli, G.; Ganau, R. Distance-Based Agglomeration Externalities and Neighbouring Firms’ Characteristics. *Reg. Stud.* **2018**, *52*, 922–933. [[CrossRef](#)]
27. Shao, S.; Tian, Z.; Yang, L. High Speed Rail and Urban Service Industry Agglomeration: Evidence from China’s Yangtze River Delta Region. *J. Transp. Geogr.* **2017**, *64*, 174–183. [[CrossRef](#)]
28. Chen, Y.; Zhu, Z.; Cheng, S. Industrial Agglomeration and Haze Pollution: Evidence from China. *Sci. Total Environ.* **2022**, *845*, 157392. [[CrossRef](#)] [[PubMed](#)]
29. Wei, W.; Zhao, L.; Liu, Z. How Does Industrial Agglomeration Affect Firms’ Energy Consumption? Empirical Evidence from China. *Indoor Built Environ.* **2023**, *32*, 1523–1536. [[CrossRef](#)]
30. Hu, S.; Song, W.; Li, C.; Zhang, C.H. The Evolution of Industrial Agglomerations and Specialization in the Yangtze River Delta from 1990–2018: An Analysis Based on Firm-Level Big Data. *Sustainability* **2019**, *11*, 5811. [[CrossRef](#)]
31. Chen, W.; Zhang, S.; Kong, D.; Zou, T.; Zhang, Y.; Cheshmehzangi, A. Spatio-Temporal Evolution and Influencing Factors of China’s ICT Service Industry. *Sci. Rep.* **2023**, *13*, 9703. [[CrossRef](#)]
32. Briant, A.; Combes, P.P.; Lafourcade, M. Dots to Boxes: Do the Size and Shape of Spatial Units Jeopardize Economic Geography Estimations? *J. Urban Econ.* **2010**, *67*, 287–302. [[CrossRef](#)]
33. Lennert, M. The Use of Exhaustive Micro-Data Firm Databases for Economic Geography: The Issues of Geocoding and Usability in the Case of the Amadeus Database. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 62–86. [[CrossRef](#)]
34. Marcon, E.; Puech, F. Evaluating the Geographic Concentration of Industries Using Distance-Based Methods. *J. Econ. Geogr.* **2003**, *3*, 409–428. [[CrossRef](#)]
35. Buzard, K.; Carlino, G.A.; Hunt, R.M.; Carr, J.K.; Smith, T.E. The Agglomeration of American R&D Labs. *J. Urban Econ.* **2017**, *101*, 14–26. [[CrossRef](#)]
36. Ripley, B.D. The Second-Order Analysis of Stationary Point Process. *J. Appl. Probab.* **1976**, *13*, 255–266. [[CrossRef](#)]
37. Besag, J.E. Comments on Ripley’s Paper. *J. R. Stat. Soc. B* **1977**, *39*, 193–195.
38. Barff, R.A. Industrial Clustering and the Organization of Production: A Point Pattern Analysis of Manufacturing in Cincinnati, Ohio. *Ann. Assoc. Am. Geogr.* **1987**, *77*, 89–103. [[CrossRef](#)]
39. Duranton, G.; Overman, H.G. Exploring the Detailed Location Patterns of U.K. Manufacturing Industries Using MicroGeographic Data. *J. Reg. Sci.* **2008**, *48*, 213–243. [[CrossRef](#)]
40. Marcon, E.; Puech, F. Measures of the Geographic Concentration of Industries: Improving Distance-Based Methods. *J. Econ. Geogr.* **2010**, *10*, 745–762. [[CrossRef](#)]
41. Lang, G.; Marcon, E.; Puech, F. Distance-Based Measures of Spatial Concentration: Introducing a Relative Density Function. *Ann. Reg. Sci.* **2020**, *64*, 243–265. [[CrossRef](#)]

42. Kukuliač, P.; Horák, J. W Function: A New Distance-Based Measure of Spatial Distribution of Economic Activities. *Geogr. Anal.* **2017**, *49*, 199–214. [[CrossRef](#)]
43. Li, J.; Zhang, W.; Yu, J.; Chen, H. Industrial Spatial Agglomeration Using Distance-Based Approach in Beijing, China. *Chin. Geogr. Sci.* **2015**, *25*, 698–712. [[CrossRef](#)]
44. Laajimi, R.; Le Gallo, J.; Benammou, S. What Geographical Concentration of Industries in the Tunisian Sahel? Empirical Evidence Using Distance-Based Measures. *Tijdschr. Econ. Soc. Geogr.* **2020**, *111*, 738–757. [[CrossRef](#)]
45. Huang, Y.; Sun, W. Spatiotemporal Change Characteristics and Differences of Manufacturing Industry Agglomeration in the Beijing-Tianjin-Hebei Region. *Prog. Geogr.* **2021**, *40*, 2011–2024. [[CrossRef](#)]
46. Yang, S.; Ma, D.; Shen, Z.; Wen, L.; Dong, L. The Impact of Artificial Intelligence Industry Agglomeration on Economic Complexity. *Ekonom. Istraz.* **2023**, *36*, 1420–1448. [[CrossRef](#)]
47. Barlet, M.; Briant, A.; Crusson, L. Location Patterns of Service Industries in France: A Distance-Based Approach. *Reg. Sci. Urban Econ.* **2013**, *43*, 338–351. [[CrossRef](#)]
48. Behrens, K.; Bougna, T. An Anatomy of the Geographical Concentration of Canadian Manufacturing Industries. *Reg. Sci. Urban Econ.* **2015**, *51*, 47–69. [[CrossRef](#)]
49. Aleksandrova, E.; Behrens, K.; Kuznetsova, M. Manufacturing (Co)Agglomeration in a Transition Country: Evidence from Russia. *J. Reg. Sci.* **2019**, *60*, 88–128. [[CrossRef](#)]
50. de Almeida, E.T.; Da Mota Silveira Neto, R.; de Moraes Rocha, R. Manufacturing Location Patterns in Brazil. *Pap. Reg. Sci.* **2022**, *101*, 839–873. [[CrossRef](#)]
51. de Almeida, E.T.; Neto, R.D.M.S.; Rocha, R.D.M. The Spatial Scope of Agglomeration Economies in Brazil. *J. Reg. Sci.* **2023**, *63*, 820–863. [[CrossRef](#)]
52. Mori, T.; Smith, T.E. A Probabilistic Modeling Approach to the Detection of Industrial Agglomerations. *J. Econ. Geogr.* **2014**, *14*, 547–588. [[CrossRef](#)]
53. Maddah, L.; Arauzo-Carod, J.-M.; López, F.A. Detection of Geographical Clustering: Cultural and Creative Industries in Barcelona. *Eur. Plan. Stud.* **2023**, *31*, 554–575. [[CrossRef](#)]
54. Yu, Z.; Zu, J.; Xu, Y.; Chen, Y.; Liu, X. Spatial and Functional Organizations of Industrial Agglomerations in China's Greater Bay Area. *Env. Plan. B-Urban Anal. City Sci.* **2022**, *49*, 1995–2010. [[CrossRef](#)]
55. Yu, Z.; Xiao, Z.; Liu, X. Characterizing the Spatial-Functional Network of Regional Industrial Agglomerations: A Data-Driven Case Study in China's Greater Bay Area. *Appl. Geogr.* **2023**, *152*, 102901. [[CrossRef](#)]
56. Lu, C.; Yu, C.; Xin, Y.; Zhang, W. Spatial Distribution Characteristics and Influencing Factors on the Retail Industry in the Central Urban Area of Lanzhou City at the Scale of Daily Living Circles. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 344. [[CrossRef](#)]
57. Carr, J.K.; Fontanella, S.A.; Tribby, C.P. Identifying American Beer Geographies: A Multiscale Core-Cluster Analysis of U.S. Breweries. *Prof. Geogr.* **2019**, *71*, 185–196. [[CrossRef](#)]
58. Buzard, K.; Carlino, G.A.; Hunt, R.M.; Carr, J.K.; Smith, T.E. Localized Knowledge Spillovers: Evidence from the Spatial Clustering of R&D Labs and Patent Citations. *Reg. Sci. Urban Econ.* **2020**, *81*, 103490. [[CrossRef](#)]
59. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*, 1st ed.; Routledge: London, UK, 2018; ISBN 978-1-315-14091-9.
60. National Public Service Platform for Standards Information. Available online: <https://openstd.samr.gov.cn/> (accessed on 2 March 2023).
61. Wu, K.; Wang, Y.; Zhang, H.; Liu, Y.; Ye, Y.; Yue, X. The Pattern, Evolution, and Mechanism of Venture Capital Flows in the Guangdong-Hong Kong-Macao Greater Bay Area, China. *J. Geogr. Sci.* **2022**, *32*, 2085–2104. [[CrossRef](#)]
62. Hui, E.C.M.; Li, X.; Chen, T.; Lang, W. Deciphering the Spatial Structure of China's Megacity Region: A New Bay Area—The Guangdong-Hong Kong-Macao Greater Bay Area in the Making. *Cities* **2020**, *105*, 102168. [[CrossRef](#)]
63. Chen, S.; Zhang, F.; Zhang, Z.; Yu, S.; Qiu, A.; Liu, S.; Zhao, X. Multi-Scale Massive Points Fast Clustering Based on Hierarchical Density Spanning Tree. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 24. [[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.