人口學刊
第 65期，2022 年 12 月，頁 1-41
Journal of Population Studies
No. 65, December 2022, pp. 1-41

Special Issue
DOI:10.6191/JPS.202212_(65).0001

A Critical Scale for Analyzing Point Patterns: An Analysis of Dengue Fever Cases in Kaohsiung City

Wei Chien Benny Chin*

* Research Fellow, Department of Geography, National University of Singapore.
E-mail: wcchin@nus.edu.sg

Received: November 15, 2022; accepted: March 30, 2023.

Abstract

In a disease outbreak context, disease cases are usually presented by using point distribution data. Due to the scale-invariant issue of point data and the scaling issue of the modifiable areal unit problem, identifying a critical scale for the analysis of point patterns, such as the clustering phenomenon, is important. This study proposes a novel data-driven framework for calculating the critical scale based on two traditional concepts: (1) the point-region quadtree spatial indexing method and (2) the box counting method for fractal pattern analysis. Both concepts capture the spatial scaling process and serve as the core concepts of the proposed framework. Using dengue fever cases in Kaohsiung City, Taiwan, during the past two decades, the critical scale was identified for each outbreak year. Two clustering analysis approaches were used to test the resulting critical scales, including kernel density estimation and density-based spatial clustering application with noise. Both clustering analyses involved distance parameter settings. Therefore, through the setting of search radii, the two clustering methods were used as a tool to explore the clustering patterns under different scale levels. In summary, the identified critical scales can better capture the spatial patterns of point data.

Keywords: spatial scaling, fractal dimension, dengue fever, clustering phenomenon, critical scale

I. Introduction

Unlike other spatial data, point data is scale invariant (Cressie 1993; Goodchild and Mark 1987). A point located in a district can also be observed as being located in a city or country. Three points located in the same city can be seen as a cluster if viewed at the country level, but the three points may be located in three different districts, creating a dispersed pattern at the district level. Similarly, when many disease cases occur in one district, these cases can be seen as one cluster if viewed at the country level, but they can also be considered dispersed if viewed from a building or village level (i.e., the cases locate in different parts of the district). The point pattern can therefore be analyzed with various scales for the same point data set. This led to this study's main research question: What is the optimal scale for analyzing point patterns?

Scaling studies provided a direction to answer this question. Previous studies suggested that some point events experienced the scale-invariant phenomenon, where the clustering pattern on a lower scale can also be observed on a higher scale (Agterberg 2013; Frankhauser 2015). These point patterns tend to be self-similar across scales. This scale-invariant phenomenon indicates that some point events or point distributions have a scaling nature, which means that the clustering pattern exists at different scales (Goodchild 2011). In simple words, several points occur near each other, forming a cluster; when zoomed out, the view reveals that many clusters occur beside each other, forming a cluster of clusters. This situation suggests that a range of scales presents a similar point distribution pattern.

However, the well-studied modifiable areal unit problem (MAUP), also called the Openshaw effect (Goodchild 2022), indicates that point

distribution suffers from a zoning and scaling problem (Gerell 2017; Openshaw 1981). The scaling problem states that when one aggregates the same set of points into different areal units (e.g., county and district level administrative boundaries), the aggregated patterns observed through the two levels of boundaries can yield different observations. While the MAUP raises a counterargument to the aforementioned scaling phenomenon, this situation also means that the range of scales has a limit—when the scale is beyond a particular level, the pattern may change.

Analyzing point scaling produced a similar finding. The analysis of the point distribution's fractal pattern (Agterberg 2013; Carlson 1991; Frankhauser 2015) suggested that a bi-fractal pattern exists in most point distributions (Agterberg 2013; Raines 2008). Indeed, when the number of occupied boxes (OB) was plotted against the box size—the standard box counting method for fractal analysis—a turning point (i.e., the roll-off effect, also called the fall-off effect) continually occurred. The existence of a turning point means that when the box becomes smaller (higher resolution) from larger (lower resolution), the number of OB increases at a constant rate (a straight and steep increasing line) before the turning point. After the turning point, the increment rate becomes a flat line, which indicates that the scaling of boxes yields a different pattern when the box continues to become smaller (higher resolution). A bi-fractal pattern is often observed in empirical studies of geographical characteristics, such as urban forms or population distribution (White and Engelen 1993; White et al. 2015), which means that the roll-off effect (i.e., the small scales where the patterns are not aligned with the global structure) was not representative of the point pattern analysis; these scales were thus usually excluded or truncated (Agterberg 2013; Pickering et al. 1995). This situation suggests that the small scales were not useful for exploring the point patterns due to their locality (only occurring in a small

part of the area) and conflict with the overall (or global) spatial pattern (Pickering et al. 1995; Walsh et al. 1991). In previous studies, the critical scale was defined as the scale's midpoint where the global pattern that represents the overall spatial structure intersects the local pattern that occurs in only a certain region (Velázquez et al. 2016; Wiegand and Moloney 2004). The roll-off effect and the existence of a turning point provide a hint for identifying the optimal scale for analyzing point distribution.

In this study, the roll-off effect was utilized to identify the critical scale of point distribution. This study used the scaling process and analysis of point distribution from the fractal pattern analysis field, scaling the study site from a low scale level (low resolution and large boxes) to a high scale level (high resolution and small boxes). The scaling process indicated that more spatial information is revealed when the scale level is increased. The critical scale is defined as the highest scale level before the turning point (i.e., the scale level containing substantial detail on spatial point distribution while not being overly detailed). In this study, a framework was proposed to investigate spatial point patterns and identify the critical scale. In the framework, the point-region quadtree (PR-Qtree) data structure is used to construct a two-dimensional point distribution quadtree structure to analyze the spatial scaling process. This study used the box counting method to conceptualize and analyze the scaling process (Agterberg 2013; Sémécurbe et al. 2016). Dengue fever cases that occurred in Kaohsiung City, Taiwan, were used as case studies to demonstrate the analysis framework.

II. Materials and Method

A framework for analyzing the point distribution's scaling pattern is proposed (Figure 1). This framework can be used to identify point patterns'

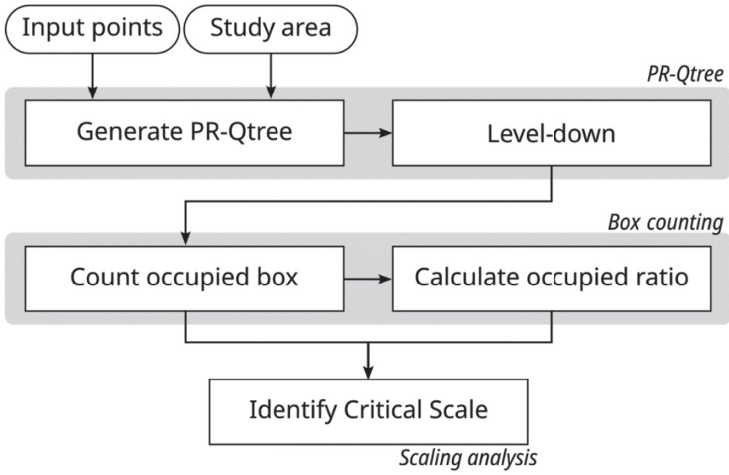

Figure 1. The calculation framework for identifying the critical scale

Note: PR-Qtree: point-region quadtree.

critical scale. The PR-Qtree data structure was used as the framework's foundation, and the box counting method—a calculation process developed in the study of fractal dimension (FD)—was applied. To use the PR-Qtree for box counting, one must execute a level-down procedure before performing the calculation. The following subsections describe the research design and experiment, followed by the procedures that reflect the proposed framework.

Research Design and Experiment

The cases of dengue fever that occurred in Kaohsiung City, Taiwan, for eight years were used to demonstrate the proposed method's application. The data was selected from 22 years (1998-2020, excluding 2000, which has no data) of dengue fever cases from Taiwan's Opendata platform (Taiwan Centers for Disease Control 2022). The original data provides the daily number of dengue fever cases by the basic statistical area, the smallest

statistical spatial unit in Taiwan. In this study, each case was reassigned to a random point location within the basic statistical area. Because dengue fever has a yearly cyclic pattern in Taiwan, the daily cases were split and aggregated according to the outbreak years, that is, from April of the outbreak year to March of the following year. A study area focusing on the core region of Kaohsiung City (east-west direction: 174 km to 187 km; north-south direction: 2,497 km to 2,510 km; projection: EPSG 3826) with a size of 13 × 13 km^2 was used throughout the study. The cases were filtered by study area, and Table 1 reveals the number of cases in the study area by outbreak year. The eight years with more than 400 cases (2002, 2006, 2009, 2010, 2011, 2012, 2014, and 2015) were selected to demonstrate the OB and ratio in the result's first section. Four years of data with different sizes, including 2011 (1,089 cases), 2002 (4,188 cases), 2014 (12,897 cases), and 2015 (16,392

Table 1. The number of dengue fever cases within the study area by outbreak year

No.	Year of outbreak	Number of cases	No.	Year of outbreak	Number of cases
1	1998	86	13	2010	966
2	1999	3	14	2011	1,089
3	2000	No data	15	2012	432
4	2001	216	16	2013	81
5	2002	4,188	17	2014	12,897
6	2003	12	18	2015	16,392
7	2004	52	19	2016	28
8	2005	105	20	2017	23
9	2006	851	21	2018	51
10	2007	151	22	2019	97
11	2008	278	23	2020	4
12	2009	620	24	2021	No data

cases), were chosen for clustering analyses. A small data set (23 cases) from outbreak year 2017 was used for the method's description below.

Three parts of the analyses were presented in the result section: (1) the OB and ratio from the calculation; (2) the kernel density estimation (KDE) analysis that used different bandwidths to demonstrate the effects of different scale levels; and (3) the density-based spatial clustering application with noise (DBSCAN) analysis that used different searching radii (epsilons) to search for neighboring events. KDE is an approach to measuring the concentration levels of point events within a specified distance based on a kernel function (Scott 1992; Silverman 1986). The kernel function is usually a Gaussian distribution (i.e., bell shape), and it is used to determine the weight of a point at a distance from each of the target cells. The distance parameter (i.e., bandwidth) is used to define the Gaussian distribution's shape. A shorter bandwidth generates a narrower kernel function and vice versa. DBSCAN is an approach for identifying individual clusters from point event data (Ester et al. 1996; Schubert et al. 2017). In the DBSCAN calculation, the distance parameter is a search radius for neighbors—other points within the distance—from each target point. With another minimum point parameter, if a point has numerous neighbors (i.e., more than the minimum point parameter), the point is identified as a core point (i.e., a high-density point). A cluster is detected as a group of connected core points through the connections between core points that fall within the distance parameter.

For KDE and DBSCAN, a distance parameter was key to determining the density or clustering relationships, which makes the two methods suitable for exploring the point patterns at different scale levels. The different distance parameter settings based on the scale levels for the same group of points were thus tested to compare the resulting density or

clustering pattern with the distance parameter set as the search radius of various scale levels. KDE (with Gaussian kernel function) and DBSCAN are performed with the Python scikit-learn package (1.2.1), and the PR-Qtree and critical scale calculations are written in the Python (3.9.15) scripting language.

PR-Qtree

The PR-Qtree (Orenstein 1982; Samet 1984) is a quadtree data structure that can be used as a spatial indexing approach to increase the performance of point and range queries. As with other two-dimensional quadtree approaches (Venables and Ripley 2002), a PR-Qtree divides the spatial area into four sub-spaces to store the points. In this study, a PR-Qtree divided a two-dimensional study space into equal-sized hierarchical quadrants to store each point in one quadrant cell. Since each depth level in a PR-Qtree indicates a specific quadrant box size, the depths can represent spatial scale levels. A PR-Qtree is therefore suitable and convenient for analyzing the points' scaling properties.

Figure 2 presents a demonstration of inserting four points into the PR-Qtree. Inserting point A (Figure 2[a]) is straightforward. Since no other points are in the tree/space, one large box covering the entire study area is generated as the root node, and the first point occupies the large box (depth-0 level; Figure 2[e]). Inserting point B (Figure 2[b]) causes the first box to contain two data points; the box is thus divided into four quadrant boxes (four branches in the tree, each represents a quadrant box), and the two points (A and B) are then moved to depth-1 level (Figure 2[f]). Similarly, when point C is inserted, it colocates with point B in the upper left box (Figure 2[c]); the box/branch is therefore split, and points B and C are moved to the depth-2 level branches (Figure 2[g]). Point D (Figure 2[d]) does not share

a box with other points; inserting point D can thus be achieved by adding a node at the empty branch (Figure 2[h]).

Level-Down

The PR-Qtree was not designed to be used as an analysis tool. To use it as the foundation for the box counting method, one needs an additional procedure to prepare the tree for the box counting calculation. The box counting of different scale levels relies on the number of nodes at each depth level. For example, in Figure 2, points A and D stop at the depth 1 level, which means that the two points would not have any records at a lower level (e.g., depth 2). This study introduced the level-down process, which pulls all points (i.e., leaf nodes, also called black nodes) to the lowest depth level. All points therefore have a parent node (i.e., gray node) on all depth levels above the leaf node, and the total count of black nodes is the total number of points.

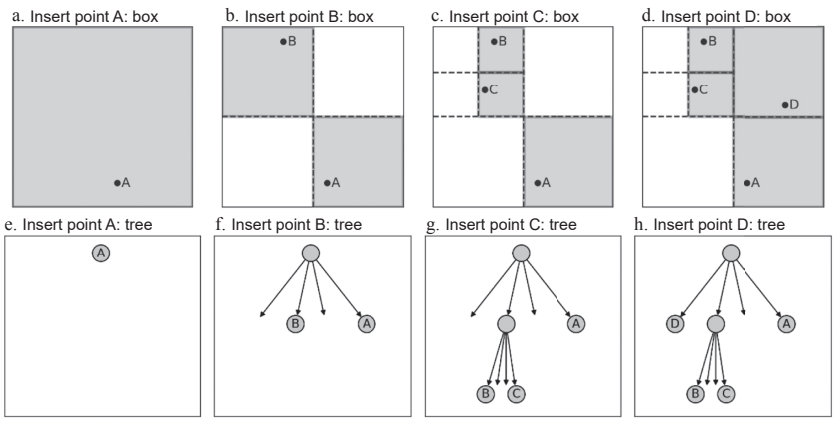

Figure 2. Example of inserting four points into a point-region quadtree (PR-Qtree)

Note: The four points were extracted from the dengue fever at Kaohsiung City in 2017 for demonstration.

Figure 3 depicts the boxes before and after the level-down process for the dengue fever cases of 2017. Altogether, 23 points were inserted. Several large OB existed before the level-down process (Figure 3[a]). Nevertheless, all OB were split to the lowest level after the level-down process (Figure 3[b]).

In summary, through the PR-Qtree conversion (including the level-down process), a point distribution can be converted into a quadtree structure that is almost unique for a specific set of points. The PR-Qtree stores the point pattern in a hierarchy structure that preserves the scaling relationships between OB of different levels—an occupied child box must have an occupied parent box, and an occupied parent box must have at least one occupied child box. The tree's depth represents different scale levels. The study area for the data sets used in this study was 13×13 km^2. Table 2 provides each scale level's corresponding box size.

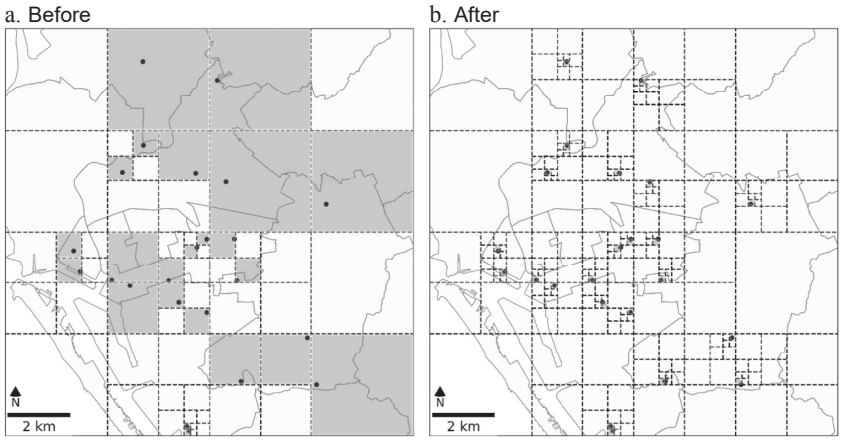

Figure 3. Comparison of the resulting boxes between before and after level-down

Table 2. The corresponding box sizes (side length) and the radius (half of the side length) of each scale level

Scale level	Size ratio	Box size (m)	Radius (m)
0	$1 : 2^0$	13,000.000000	6,500.000000
1	$1 : 2^1$	6,500.000000	3,250.000000
2	$1 : 2^2$	3,250.000000	1,625.000000
3	$1 : 2^3$	1,625.000000	812.500000
4	$1 : 2^4$	812.500000	406.250000
5	$1 : 2^5$	406.250000	203.125000
6	$1 : 2^6$	203.125000	101.562500
7	$1 : 2^7$	101.562500	50.781250
8	$1 : 2^8$	50.781250	25.390625
9	$1 : 2^9$	25.390625	12.695312
10	$1 : 2^{10}$	12.695312	6.347656
11	$1 : 2^{11}$	6.347656	3.173828
12	$1 : 2^{12}$	3.173828	1.586914
13	$1 : 2^{13}$	1.586914	0.793457
14	$1 : 2^{14}$	0.793457	0.396729

The FD and the Box Counting Method

To empirically measure the fractal geometry of geographical features in the real world, researchers have introduced the box counting method and applied it to geographical studies (Caballero et al. 2022; Feng and Chen 2010; Frankhauser 2015; Jiang and Liu 2012; Wu et al. 2020). In simple words, the box counting method measures fractal geometry by quantifying the OB on different scale levels. For example, Figure 4 presents the counting of OB at the six levels. Level 0 has only one box, and all points fall inside it; the count is thus $OB = 1$. At level 1, the previous box was divided into four quadrants, and all four boxes contained some points; the count is therefore $OB = 4$. For level 2, five out of 16 boxes were empty; the count

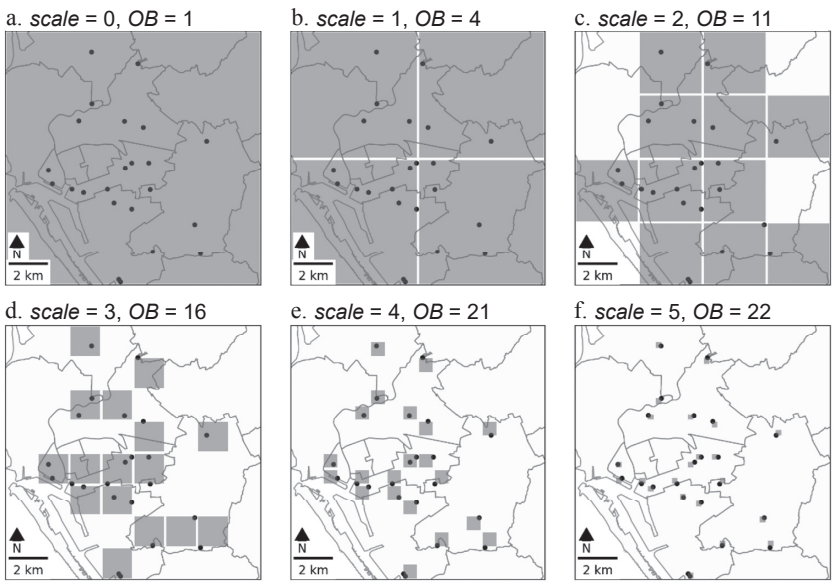

Figure 4. The counting of occupied boxes (OB) in the first 6 scale levels

is thus *OB* = 11. Similarly, for levels 3-5, the shaded boxes highlighted the points' locations, and the counted boxes totaled 16, 21, and 22, respectively. Level 5 has only one box containing more than one point. The total number of points in this data set is 23, and the lowest depth is level 6, where the last two points were split into independent boxes.

With the traditional box counting method using box sizes (*l*, side length) and counts, the point distribution's *FD* can be calculated by fitting the slope (Equation [1]) of the line in the box size ($\log_2 l$, horizontal axis) to the box count ($\log_2 OB$, vertical axis) plots (Frankhauser 2004, 2015). In a standard box counting calculation, the horizontal axis is the box length; when the box's length increases, the box becomes larger, and the number of *OB* decreases. The slope of the line is therefore expected to be negative, and the *FD* is the negative of the slope value (conversion to positive).

$$FD = -\frac{\log_2 OB}{\log_2 l} \quad \text{.....(1)}$$

Again, the tree's depth captures the concept of scale level (*scale* = {1, 2, 3, …}) and each scale level corresponds to a specific box size (or side length, l_{scale}; Table 2). The FD calculation can be simplified with the PR-Qtree. The side length of each cell at a depth level can be calculated as the result of dividing the entire side length (L) by 2 with an exponent of the scale level, that is, $l_{scale} = L/2^{scale}$ (see Appendix 1). Equation (2) could thus be used to convert the box size (side length) to scale level in this study.

$$scale = -\log_2 \frac{l_{scale}}{L} \quad \text{.....(2)}$$

Based on Equations (1) and (2), the *FD* can be simplified as illustrated in Equation (3) (see Appendix 2) using the scale levels and corresponding occupied box count (OB_{scale}). No negative sign occurs in the equation, which indicates that the two variables' slope is positive (i.e., the higher the scale level, the higher the log number of *OB*). This change occurs because the scale level starts from the largest box (*scale* = 0) to smaller boxes, and the number of *OB* increases in this scaling process. Because the calculation in the following section is similar to the FD calculation, in this paper, the FD is denoted as the S_{OB} (i.e., slope calculated from the OB plot) since it is calculated based on OB_{scale}.

$$FD = \frac{\log_2 OB_{scale}}{scale} = S_{OB} \quad \text{.....(3)}$$

The Identification of Critical Scale

Previous studies have indicated that the distributions of geographical

point events reveal a bi-fractal pattern (Agterberg 2013; Chen and Wang 2013; Tannier and Pumain 2005; Thomas et al. 2008) due to the geographical features' scaling nature (Batty 2008; Goodchild and Mark 1987). This phenomenon leads to the roll-off effect, where a turning point of scale exists between the scale ranges (Agterberg 2013; Pickering et al. 1995; Walsh et al. 1991). Conceptually speaking, the turning point of scales indicates that the range of scales before the turning point experiences self-similar properties across scales—the global fractal pattern. The range of scales after the turning point experiences the roll-off effect, which means that the distribution surpasses the explanation of the global fractal pattern—a local fractal pattern that occurs in only some parts of the study area (Velázquez et al. 2016; Wiegand and Moloney 2004). In this study, the turning point of scales is called the critical scale because it represents the optimized and finest scale level that the global fractal pattern can explain and starts to turn into the local fractal pattern.

Equation (4) introduces an index, the occupied ratio (OR_{scale}), to calculate the critical scale. The OR_{scale} converts the OB_{scale} into a ratio by dividing the OB_{scale} by the total number of boxes at each level. The total number of boxes (occupied and unoccupied) equals 4 with an exponent of the scale level. Dividing the OB_{scale} by the total number of cells at each depth converts the value into a unit interval that represents the coverage ratio (i.e., OR_{scale}).

$$OR_{scale} = \frac{OB_{scale}}{4^{scale}} \quad \text{.....(4)}$$

While the OB_{scale} emphasizes the increment of the number of boxes at the earlier stage (i.e., lower scale levels), the OR_{scale} emphasizes the changes of the OB to the total box counts at the later stage (i.e., higher scale levels),

where the increment of OB is slow (i.e., the reasons why the scale levels experience the roll-off effect). In the roll-effect stage, increasing a level creates a large number of total boxes, but the increment of the OB_{scale} is relatively slow, which leads to a steep decreasing trend for the OR_{scale}. The log OR_{scale} should thus be negatively correlated to the scale levels. Similar to the S_{OB} that is fitted from the OB_{scale}, here, the local fractal pattern (S_{OR}) can be calculated based on the OR_{scale} (Equation [5]).

$$S_{OR} = -\frac{\log_2 OR_{scale}}{scale} \quad \text{.....(5)}$$

S_{OB} and S_{OR} can be calculated only when a cut-off point is determined. The cut-off point can be any scale level between the minimum and maximum scale levels (0 < cut-off level < last scale level). The scale levels before the cut-off and the corresponding OB_{scale} are used to calculate the S_{OB}, while the scale levels after the cut-off and the corresponding OR_{scale} are used to calculate the S_{OR} (Figure 5). An optimization model is used to identify the scales' optimal cut-off point (i.e., turning point) so that the OB's slope and the ratio have the highest fitness. The fitness function (see Figure 5[c] in the optimization model) is the multiplication of the correlation coefficient between the fitted line (calculated based on S_{OB} and S_{OR}) and the data points (i.e., OB_{scale} and OR_{scale}). The model attempts to fit the OB_{scale} line (Equation [3]; Figure 5[a]) and the OR_{scale} line (Equation [5]; Figure 5[b]) using the same cut-off, which can be any of the scale levels in the range. The dashed blue and red lines in Figures 5(a) and 5(b), respectively, indicate the fitted line. The Pearson correlation coefficients between fitted lines and data for global and local patterns were calculated (the blue and red lines in Figure 5[c]) and multiplied to compute fitness for each level in the range (the black

line in Figure 5[c]). The model attempts to slide the cut-off value (i.e., the turning point of scales) in OB_{scale} and OR_{scale} plots, calculate the fitting lines, and then maximize the two correlation results concurrently.

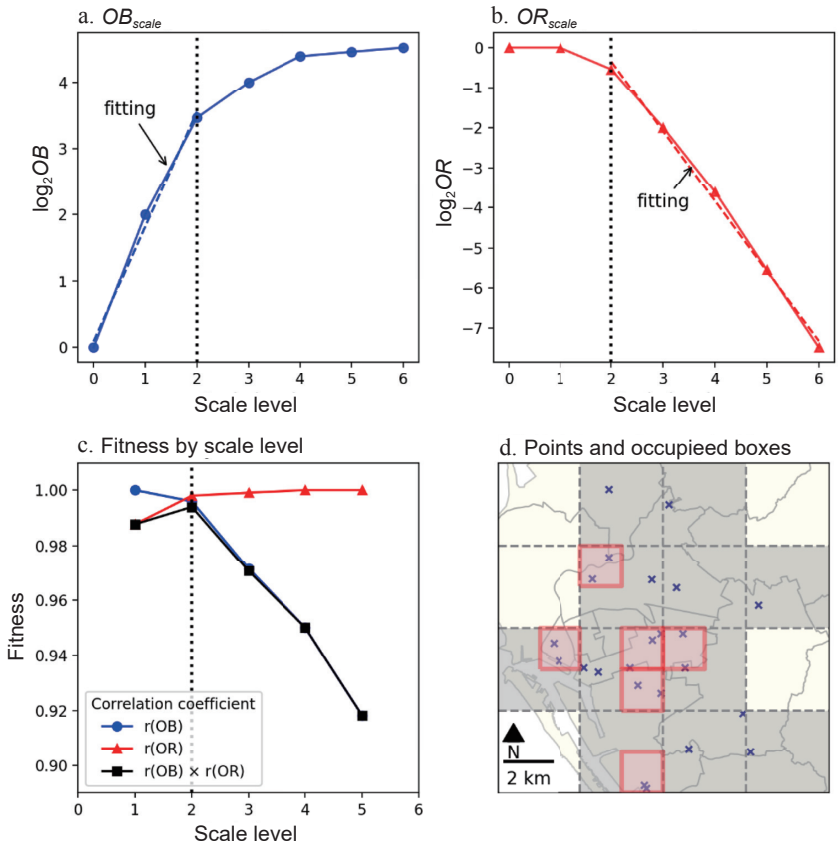

Figure 5. The calculation of the critical scales using the (a) occupied box count (OB_{scale}) plot and (b) occupied ratio (OR_{scale}) plot. (c) The level-2 is identified as the scale level with the highest fitness value. (d) The resulting optimal scale level (gray boxes) and the areas that contain multiple points (red boxes)

Note: See the online version (https://doi.org/10.6191/JPS) for the full-colored figure.

III. Results

Critical Scales of the Outbreaks in Various Outbreak Years

The critical scales for each outbreak year were calculated and are disclosed below (Table 3). The S_{OB} and S_{OR} were meant to be between 1 and 2. The critical scales ranged from 1 to 8 and depended on the number (N) of points (i.e., the larger the number of points, the higher the critical scales). This result aligned with the expectation that when the point event count is large, the scale level must be higher (higher resolution) to explore and describe the spatial pattern.

According to Table 3, the outbreak years with more cases (e.g., $N \geq 400$) tend to have a critical scale between 4 and 8. The critical scales seem to follow a semi-log relationship, log(N), with the sample size (N) because it is derived using the PR-Qtree approach. A tree-based data structure tends to have a depth of log(N). Although the critical scale did not reach the lowest depth, it reveals a semi-log relationship.

Figure 6 depicts the OB and occupied ratio plots for the eight years with $N \geq 400$. Complementing what Figure 5 demonstrates, the OB increased as a straight line from the beginning, and the occupied ratio decreased as a straight line at the end. The critical scale was a turning point after the OB_{scale}'s straight line and before the OR_{scale}'s straight line started. Since Figure 5 presents a demonstration using data from outbreak year 2017 (23 points), the OB_{scale} and the OR_{scale} stopped at scale level 6 (i.e., the last scale level). With the number of points exceeding 400, the resulting lines in Figure 6 present a similar structure; however, the last scale levels were larger than 10 (e.g., outbreak

Table 3. The slopes of the occupied box (S_{OB}) and occupied ratio (S_{OR}) lines, the critical scale, and corresponding box size (length of each side) for the 22 years of data

No.	Year	N	S_{OB}	S_{OR}	Critical scale	Box size (m)
1	1998	86	1.85	1.67	2	3,250.00000
2	1999	3	1.00	2.00	1	6,500.00000
3	2001	216	1.79	1.58	2	3,250.00000
4	2002	4,188	1.69	1.86	6	203.12500
5	2003	12	1.58	1.81	1	6,500.00000
6	2004	52	1.73	1.71	2	3,250.00000
7	2005	105	1.79	1.64	2	3,250.00000
8	2006	851	1.60	1.78	5	406.25000
9	2007	151	1.85	1.70	2	3,250.00000
10	2008	278	1.95	1.62	2	3,250.00000
11	2009	620	1.70	1.72	4	812.50000
12	2010	966	1.61	1.80	5	406.25000
13	2011	1,089	1.64	1.80	5	406.25000
14	2012	432	1.64	1.70	4	812.50000
15	2013	81	1.95	1.69	2	3,250.00000
16	2014	12,897	1.75	1.86	7	101.56250
17	2015	16,392	1.69	1.90	8	50.78125
18	2016	28	2.00	1.57	1	6,500.00000
19	2017	23	1.73	1.74	2	3,250.00000
20	2018	51	1.59	1.85	3	1,625.00000
21	2019	97	1.90	1.70	2	3,250.00000
22	2020	4	1.00	2.00	1	6,500.00000

years 2014 and 2015 were 15 and 14, respectively), which indicates that while the calculations could be scaled up to a large number of points, the patterns remained stable, and the critical scales scaled up depended on the semi-log relationships with the number of points.

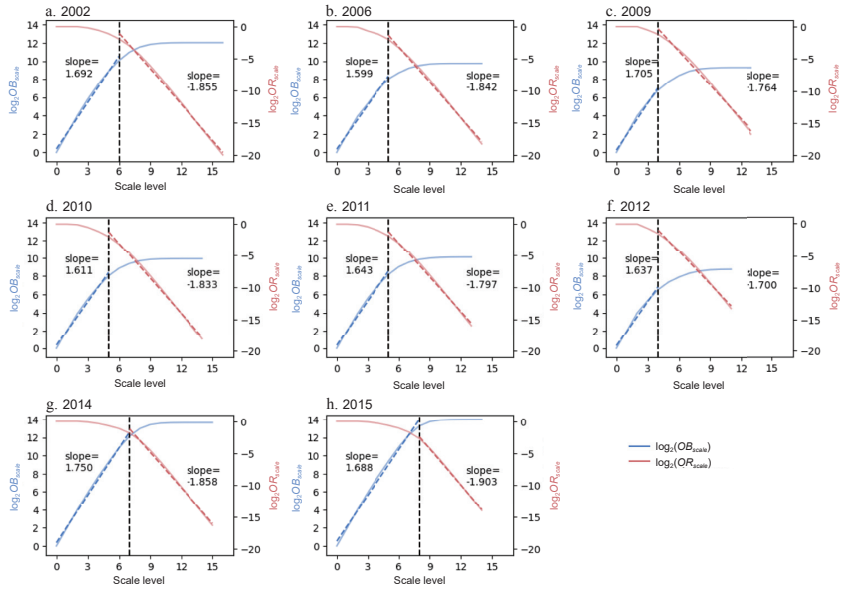

Figure 6. The occupied box (OB_{scale}) (increasing) and occupied ratio (OR_{scale}) (decreasing) plots for the eight selected years

Note: See the online version (https://doi.org/10.6191/JPS) for the full-colored figure.

The KDE Based on Varying Bandwidth

To test the results, the KDE (grid cell size = 50 m) using the different bandwidths for the four selected outbreak years: 2011 (1,089 cases), 2002 (4,188 cases), 2014 (12,897 cases), and 2015 (16,392 cases), were calculated. The corresponding critical scales for the four years (2011, 2002, 2014, and 2015) were 5, 6, 7, and 8, respectively; we therefore presented the four sets of KDE results, including the four bandwidth settings for the four scale levels (5-8) as selected scale levels, to compare the spatial patterns based on the four bandwidths. Because the bandwidth setting for the KDE calculation was a search radius that formed a circle surrounding the centers of each cell and measured the weighted point numbers that fell within the

circle, the bandwidth was set as half of the box length (Table 2) for the four selected scale levels.

Figure 7 presents the four sets of KDE (columns) for the four outbreak years (rows). For outbreak year 2011 (Figure 7, top row), the critical scale

Figure 7. The result of kernel density estimation (KDE) using bandwidth settings of four scale levels (5-8) on the four selected years (2011, 2002, 2014, and 2015)

Note: The bandwidth (radius) was set to be half of the box length of the corresponding scale levels. The boxes indicate the range of the map in the next figure.

was 5. In the comparison between the four bandwidth settings, the first map (Figure 7[a]; *scale* = 5, bandwidth = 203.1250 m) revealed a suitable pattern for observations since it indicated a few clusters with a clear pattern on the map, while the second map (Figure 7[b]; *scale* = 6, bandwidth = 101.5625 m) captured only the largest cluster on the previous map, which was presented as a small and isolated cluster; the other groups of points were presented mostly as dispersed and scattered points. At scale levels 7 and 8 (Figures 7[c] and 7[d]), no clusters were identified. For outbreak year 2002 (Figure 7, second row), the first map (Figure 7[e]; *scale* = 5, bandwidth = 203.125 m) reveals about three clusters. The second map (Figure 7[f]; *scale* = 6, bandwidth = 101.5625 m) presented more details on the largest cluster, and the two smaller clusters became small and isolated clusters as in the previous year (Figure 7[b]). No cluster can be found on the other two maps (scale levels 7 and 8; Figures 7[g] and 7[h]). The critical scale for outbreak year 2002 was level 6 (bandwidth = 101.5625 m). The results from the two years indicated that with several points between 1,000 and 5,000, the bandwidth settings of levels 5 and 6 (203.1250 m and 101.5625 m, respectively) can present sufficient details of the spatial distributions and thus be used as observation scales for the point data in the two outbreak years.

The last two rows of Figure 7 reveal the KDE of outbreak years 2014 and 2015, which contained numerous cases (12,000-16,000) compared to those in the first 2 rows. The first column (Figures 7[i] and 7[m]; *scale* = 5, bandwidth = 203.1250 m) indicates that several clusters appeared in different parts of the study area. These clusters merged, making them difficult to differentiate. These points were indeed seen as a large cluster if viewed from a higher level (e.g., the city scale). Nevertheless, with the study area extent that reveals only the fixed region, the analysis aimed to identify the spatial patterns within the study area (i.e., the local pattern). Depicting them as a

large cluster covering almost the entire study area did not help describe the pattern. The second column (Figures 7[j] and 7[n]; *scale* = 6, bandwidth = 101.5625 m) presents some spatial details, but some of these clusters were still merging. For the last two columns (Figures 7[k]-7[l] and 7[o]-7[p]), while some patterns were displayed, the main patterns were difficult to observe with the entire study area and the current cell sizes (50 m). A smaller region (black boxes) was thus highlighted and is presented in Figure 8.

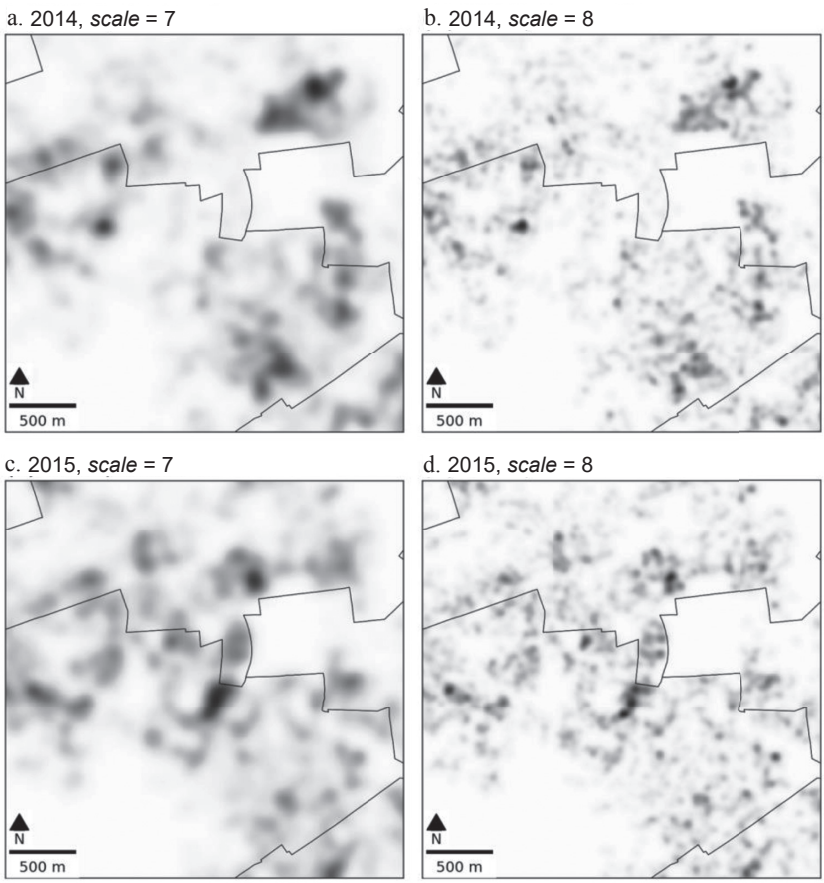

Figure 8. The zoom-in views of kernel density estimation (KDE) results for the year 2014-2015, based on scale levels 7-8

Figure 8 demonstrates the smaller region (the black box in Figure 7) KDE (grid cell size = 10 m) of outbreak years 2014 and 2015 using the bandwidths of the last two scale levels (*scale* = 7, bandwidth = 50.781250 m and *scale* = 8, bandwidth = 25.390625 m). Based on the calculation, the critical scale for 2014 was 7, and for 2015, it was 8. When the researchers used a finer resolution and smaller bandwidth, the resulting KDE presented more details of the spatial distribution over the two years. The KDE with *scale* = 8 for outbreak year 2014 seems slightly scattered, and the western group of clusters became less concentrated. For outbreak year 2015, the spatial distribution provides even more spatial details with *scale* = 8 than the previous scale level (*scale* = 7). Compared to the lower scale levels (*scale* = 5 and *scale* = 6), scale levels 7 and 8 were more suitable.

In summary, the KDE observations indicated that the suitable KDE bandwidth tends to be similar to and near the critical scale result. We therefore argue that critical scales can provide a better observation of point patterns. Since the critical scale is a scale of analysis, it indicates a range of distance (i.e., the distance from critical scale − 1 to critical scale + 1), which can be effective for the parameter settings of the search radius distance in a distance-based density estimation method. However, this study cannot identify a specific distance value for the search radius setting because the scaling process indicates a division of the length by 2, which declines fast.

The DBSCAN Based on Different Epsilons

In this study, DBSCAN was used to compare the effects of different distance parameters (epsilon) on the point distribution patterns. As described in the previous section, the data from the four outbreak years—2011 (1,089 cases), 2002 (4,188 cases), 2014 (12,897 cases), and 2015 (16,392 cases)—was used for the comparison of the four distance parameter settings (based

on scale levels 5-8, distance parameters were 203.125000 m, 101.562500 m, 50.781250 m, and 25.390625 m, respectively). Since the purpose was to compare the distance parameters, the minimum point parameter was fixed at 3. Figure 9 presents the results.

For outbreak year 2011 (Figure 9, top row), some small clusters were identified at scale level 5 (Figure 9[a]). At scale level 6 (Figure 9[b]), the clusters were isolated, and various noises (not core points) appeared. At levels 7 and 8 (Figures 9[c] and 9[d]), almost all points were identified as noises. For outbreak year 2002 (Figure 9, second row), some large clusters were detected at level 5 (Figure 9[e]), where some were great (in the southern central area) and contained some low-density regions. At scale level 6 (Figure 9[f]), the great cluster was split into several smaller clusters, one of which captured a large high-density region. At scale level 7 (Figure 9[g]), the densely populated cluster remained a cluster, but many other points were identified as noises. At scale level 8 (Figure 9[h]), almost all points were noises. The critical scales were calculated as levels 5 (2011) and 6 (2002). The observations from the DBSCAN results aligned well with the critical scale findings.

For outbreak years 2014 and 2015 (third and last rows), levels 5 and 6 (Figures 9[i]-[j] and 9[m]-[n]) presented some great clusters because these two outbreak years contained numerous case points (around 12,000 and 16,000) in the study area. Using a relatively large search radius and the same minimum number of point parameters results in many of them being connected through core points. Compared to 2011 and 2002, which contained around 1,000 and 5,000 points, the same distance parameters of scale levels 5 and 6 did not seem to work in the situations in outbreak years 2014 and 2015. At levels 7 and 8 (Figures 9[k]-[l] and 9[o]-9[p]), several clusters could be observed with the entire study area region, but the detail

differences and patterns were not visible. A smaller region (the black box area) in the city's central region was thus zoomed in in Figure 10 to observe the result.

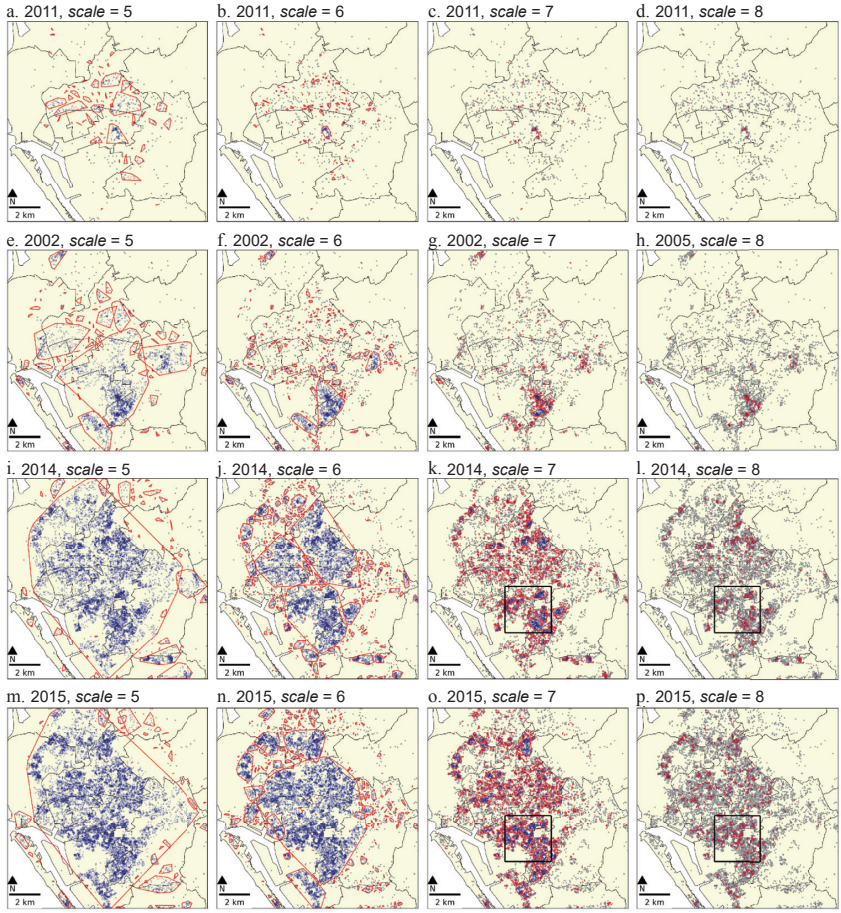

Figure 9. The clusters (red polygons) identified with density-based spatial clustering application with noise (DBSCAN) using epsilon settings of 4 scale levels (5-8) on the four selected years (2011, 2002, 2014, and 2015)

Note: 1. The square boxes indicate the range of the map in the next figure.

2. See the online version (https://doi.org/10.6191/JPS) for the full-colored figure.

The cluster map of outbreak year 2014 at scale level 7 (Figure 10[a]) reveals some large clusters together with some smaller ones. At level 8 (Figure 10[b]), several point events became noises, and the clusters were mostly isolated. For outbreak year 2015, more large clusters were found at level 7 (Figure 10[c]). At level 8 (Figure 10[d]), some identifiable clusters

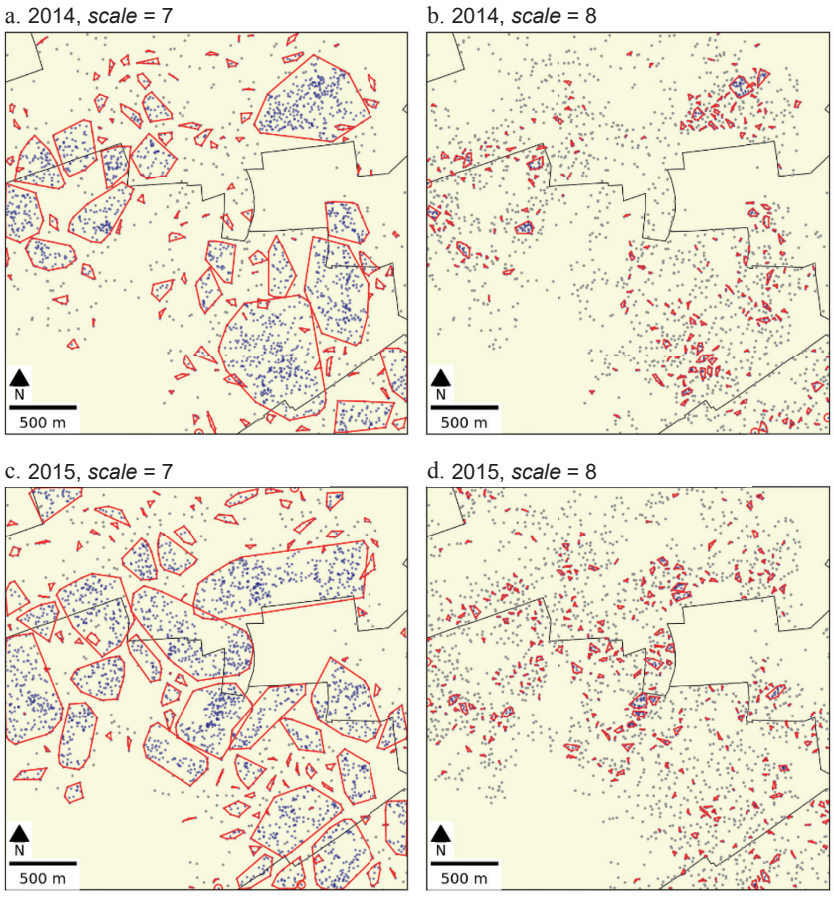

Figure 10. The zoom-in view of density-based spatial clustering application with noise (DBSCAN) results for the year 2014-2015, based on scale levels 7-8

Note: See the online version (https://doi.org/10.6191/JPS) for the full-colored figure.

remained, and the number of noises seemed less than in 2014. Level 7 thus provided a better observation of the point patterns for outbreak year 2014, whereas for outbreak year 2015, levels 7 and 8 could be suitable for exploring the clustering patterns.

According to the DBSCAN analysis, the comparisons and exploration indicated that the distance parameter settings based on the critical scales were potential and suitable options for identifying the cluster patterns. Although the presented method cannot be used to determine a specific value for the distance parameter, it can help narrow the distance range based on a range near the critical scale.

IV. Discussions

In the context of disease outbreaks, when a small number of cases occurs and these cases are distributed in different but neighboring districts, they are usually considered a cluster—the related cases. However, when many cases are scattered over the area, one is expected to be more detailed in the exploration and description of the point distribution, which necessitates using a finer scale of observation; indeed, grouping a large number of cases into a single cluster would not help describe the spatial distribution or provide insights for the community of stakeholders (e.g., the public, decision-makers, or practitioners). One must therefore adjust the analysis scales to analyze point events, which highlights the importance of identifying the critical scales from the point patterns.

This study introduced a framework for analyzing point distribution to calculate the critical scale. Using the PR-Qtree and the roll-off phenomenon in the point scaling process, the study determined that the scale level before the roll-off effect can capture point distribution's comprehensive structure

without being excessively fine, which may result in a dispersed and isolated pattern. This study used KDE and DBSCAN, two clustering pattern analysis methods, to demonstrate the changing scale levels' effect. The two methods take a search radius parameter (i.e., the bandwidth in KDE and the epsilon in DBSCAN), calculate the density, and identify clusters using the search radius as a key parameter. The two methods can thus provide a visual assessment of the spatial patterns at varying scale levels. Based on the KDE and DBSCAN analyses, the results suggested that one level above or below the critical scale (±1) is generally sufficient to present the point patterns, while the critical scale provides a better structure for the cluster patterns. The results also suggested that the scale levels beyond the ±1 range tended to provide a non-effective observation of the point distribution.

This study used the fractal pattern perspective to establish a systematic method for point pattern observation and discussion. The presented analysis used the dengue fever cases in Kaohsiung City to demonstrate point patterns' various conditions. For example, in outbreak years 2011 and 2015, 1,089 and 16,392 cases were located in the study area. The analysis scale of outbreak year 2011 was lower because the points were more scattered compared to outbreak year 2015, where the points were closer to each other. This difference means that in the case of outbreak year 2015, a coarser analysis scale may bring useless results and observations, whereas in outbreak year 2011, a finer analysis scale would not bring more information on the spatial pattern or even identify most points as noises. The visual observation from KDE and DBSCAN highlighted the condition that when fewer points occur, the spatial patterns that are calculated with a larger search radius may offer a better pattern, while a shorter radius may identify many isolated points. In contrast, when the points are dense and the number of points is large, a shorter radius would identify more local clusters

(non-isolated points), revealing more details on the spatial pattern, while a large radius would connect several points as some large clusters. This study therefore used the scaling process concept to identify a better viewing scale for point distributions.

In this study, the PR-Qtree was used as the main tool to construct points' spatial indexes and represent the spatial scaling process, that is, from the coarsest resolution with only one root node (box) representing all data points to the finest resolution where each data point is separated in an independent box. This spatial scaling process facilitated calculating the OB and occupied ratio at different scale levels (i.e., the PR-Qtree depth). The spatial scaling process based on the PR-Qtree divided the space into four equal-sized quadrants, each containing a side length of half of the previous box's side length, which indicated a logarithmic relationship of the scale level, as described in Equation (2). In other words, a log (with a base of 2) relationship exists because the length size (Table 2) is divided by 2 while the scale level is increased by 1. Although changes may be large on the linear scale, it follows the traditional box counting method concept (Agterberg 2013; Raines 2008) that both axes (the horizontal axis as the length of an individual box and the vertical axis as the number of OB) should be converted to a log-log scale. The PR-Qtree's tree structure was thus suitable for analyzing the spatial scaling process.

In many studies, search radius selection is a key issue in the analysis and depends mostly on domain knowledge or some traditional bandwidth selection approaches. For example, disease clusters are commonly defined as two or more cases (residential or working locations) occurring within 150 m or 250 m; for human activity-related studies, the distance is usually set as the walkable distance (around 400 m). The search radius can also be determined by analyzing the point pattern itself, that is, the Silverman

(1986) and Scott (1992) methods for KDE and the elbow (knee) of the *k*-nearest neighbor plot method for DBSCAN (Ester et al. 1996; Schubert et al. 2017). Nevertheless, neither the domain knowledge approach nor the bandwidth selection approach can capture the overall spatial pattern and eliminate the situation of an overly coarse or fine scale of analysis. The domain knowledge approach focused on the event, whereas the bandwidth selection approach focused on the distance between each pair of points and emphasized the statistical characteristics of the distances (i.e., second-order analysis). These approaches do not aim to analyze the point pattern's scaling structure. This study thus presented an analysis method for analyzing the scaling process and identifying the critical scale from point distributions.

Since the proposed framework was based on the box counting method, it also inherited some issues from this concept, including the influences of the study area's size and the locations of the study area's four corners. The size of the entire study area determined the sizes of different scale levels. The study area was repeatedly divided into equal-sized quadrants, which means that the size of the entire study area influenced the boxes at all levels. Similarly, the location of the study area's four corners affects the split of the initial four quadrants (i.e., the location of the study area's center point), which affects the four quadrants' center points. This iterative splitting process then depends on selecting the four corners' coordinates. These points and the study area size therefore also influence calculating the *OB*. However, because the PR-Qtree data structure reduced the size of the quadrants and the box length by dividing them by 2, the OB_{scale} were in log scale. This division accelerates the speeds of increasing resolutions, which reduces the influences of the study area size and four corners' coordinates through the quadrant splits and OB_{scale}. In most cases, the influences of the box issues may result in the critical scale shifting by one (±1).

V. Conclusion

This study aimed to provide a framework for calculating critical scales, which can be used to analyze point patterns. Various events in population or urban studies were recorded in point coordinate data (e.g., diseases, crimes, point of interests, and geotagged social media posts). The framework proposed in this study can be applied to those data sets to identify the critical scale of analysis. From the technical perspective, the critical scale is calculated based on the spatial scaling process, which is based on calculating the FDs (i.e., the box counting method). From the application perspective, it captures the natural and intuitive choices for identifying clusters. The natural selection of scales for identifying clusters can result from the FD analysis that aims to reveal the complexity of a pattern—in this case, the complexity of point patterns. When the disease case count is low, the overall pattern is less complex, and lower scale levels can thus be used to describe the spatial distribution. When many cases occur, the overall pattern may become more complex, which necessitates using a higher resolution to analyze and discuss the point pattern. One of this study's key contributions is therefore the framework for identifying the critical scale.

The following paragraph describes future directions for developing this framework. First, more tests on the influences of the study area size and corners are needed; although the influences are addressed theoretically above, more statistical tests and analyses are required. Second, while the critical scale suggested the scale for observation, further calculating a more precise distance parameter around the critical scales is possible. Third, the scales for observation concept does not have a commonly used index, which makes validating the result difficult. Future researchers could develop a

metric for exploring along the line. Fourth, the framework using the PR-Qtree and box counting methods could be further discussed and tested with the complete spatial randomness concept (Caballero et al. 2022). Fifth, more theoretical and empirical spatial scaling analysis could be discussed (Batty 2008; Bettencourt et al. 2007). Sixth, the analysis process can only consider the location of points (i.e., non-weighted point distribution). This study did not consider points with weights (e.g., population, capacity of points, and number of cases). Future analysis could expand on the current framework for the weighted point data. These future studies could further enhance the comprehensiveness of the analysis of point patterns' complexity.

References

Agterberg, F. P. 2013. “Fractals and Spatial Statistics of Point Patterns.” *Journal of Earth Science* 24(1): 1-11. doi:10.1007/s12583-013-0305-6

Batty, M. 2008. “The Size, Scale, and Shape of Cities.” *Science* 319(5864): 769-771. doi:10.1126/science.1151419

Bettencourt, L. M. A., J. Lobo, D. Helbing, C. Kühnert, and G. B. West. 2007. “Growth, Innovation, Scaling, and the Pace of Life in Cities.” *Proceedings of the National Academy of Sciences* 104(17): 7301-7306. doi:10.1073/pnas.0610172104

Caballero, Y., R. Giraldo, and J. Mateu. 2022. “A Spatial Randomness Test Based on the Box-Counting Dimension.” *AStA Advances in Statistical Analysis* 106: 499-524. doi:10.1007/s10182-021-00434-4

Carlson, C. A. 1991. “Spatial Distribution of Ore Deposits.” *Geology* 19(2): 111-114. doi:10.1130/0091-7613(1991)019<0111:SDOOD>2.3.CO;2

Chen, Y. and J. Wang. 2013. “Multifractal Characterization of Urban Form and Growth: The Case of Beijing.” *Environment and Planning B: Planning and Design* 40(5): 884-904. doi:10.1068/b36155

Cressie, N. A. C. 1993. “Spatial Point Patterns.” Pp. 577-724 in *Statistics for Spatial Data*, Rev. ed., edited by N. A. C. Cressie. New York, NY: Wiley. doi:10.1002/9781119115151.ch8

Ester, M., H. P. Kriegel, J. Sander, and X. Xu. 1996. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.” Pp. 226-231 in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, edited by E. Simoudis, J. Han, and U. Fayyad. Palo Alto, CA: AAAI Press.

Feng, J. and Y. Chen. 2010. "Spatiotemporal Evolution of Urban Form and Land-Use Structure in Hangzhou, China: Evidence from Fractals." *Environment and Planning B: Planning and Design* 37(5): 838-856. doi:10.1068/b35078

Frankhauser, P. 2004. "Comparing the Morphology of Urban Patterns in Europe: A Fractal Approach." Pp. 79-105 in *European Cities: Insights on Outskirts—Structures*, edited by A. Borsdorf and P. Zembri. Brussels, Belgium: COST.

Frankhauser, P. 2015. "From Fractal Urban Pattern Analysis to Fractal Urban Planning Concepts." Pp. 13-48 in *Computational Approaches for Urban Environments*, edited by M. Helbich, J. J. Arsanjani, and M. Leitner. Cham, Switzerland: Springer. doi:10.1007/978-3-319-11469-9_2

Gerell, M. 2017. "Smallest is Better? The Spatial Distribution of Arson and the Modifiable Areal Unit Problem." *Journal of Quantitative Criminology* 33(2): 293-318. doi:10.1007/s10940-016-9297-6

Goodchild, M. F. 2011. "Scale in GIS: An Overview." *Geomorphology* 130(1-2): 5-9. doi:10.1016/j.geomorph.2010.10.004

Goodchild, M. F. 2022. "The Openshaw Effect." *International Journal of Geographical Information Science* 36(9): 1697-1698. doi:10.1080/13658816.2022.2102637

Goodchild, M. F. and D. M. Mark. 1987. "The Fractal Nature of Geographic Phenomena." *Annals of the Association of American Geographers* 77(2): 265-278. doi:10.1111/j.1467-8306.1987.tb00158.x

Jiang, S. and D. Liu. 2012. "Box-Counting Dimension of Fractal Urban Form: Stability Issues and Measurement Design." *International Journal of Artificial Life Research* 3(3): 41-63. doi:10.4018/jalr.2012070104

Openshaw, S. 1981. “The Modifiable Areal Unit Problem.” Pp. 60-69 in *Quantitative Geography: A British View*, edited by N. Wrigley and R. J. Bennett. London, UK: Routledge & Kegan Paul.

Orenstein, J. A. 1982. “Multidimensional Tries Used for Associative Searching.” *Information Processing Letters* 14(4): 150-157. doi:10.1016/0020-0190(82)90027-8

Pickering, G., J. M. Bull, and D. J. Sanderson. 1995. “Sampling Power-Law Distributions.” *Tectonophysics* 248(1–2): 1-20. doi:10.1016/0040-1951(95)00030-Q

Raines, G. L. 2008. “Are Fractal Dimensions of the Spatial Distribution of Mineral Deposits Meaningful?” *Natural Resources Research* 17(2): 87-97. doi:10.1007/s11053-008-9067-8

Samet, H. 1984. “The Quadtree and Related Hierarchical Data Structures.” *ACM Computing Surveys* 16(2): 187-260. doi:10.1145/356924.356930

Schubert, E., J. Sander, M. Ester, H. P. Kriegel, and X. Xu. 2017. “DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN.” *ACM Transactions on Database Systems*, 42(3): 1-21. doi:10.1145/3068335

Scott, D. W. 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York, NY: Wiley.

Sémécurbe, F., C. Tannier, and S. G. Roux. 2016. “Spatial Distribution of Human Population in France: Exploring the Modifiable Areal Unit Problem Using Multifractal Analysis.” *Geographical Analysis* 48(3): 292-313. doi:10.1111/gean.12099

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. London, UK: Chapman and Hall/CRC.

Taiwan Centers for Disease Control. 2022. “Dengue Daily Confirmed Cases

since 1998" [Data set]. https://data.gov.tw/dataset/73178 (Date visited: September 30, 2022).

Tannier, C. and D. Pumain. 2005. "Fractals in Urban Geography: A Theoretical Outline and an Empirical Example." *Cybergeo: European Journal of Geography* 307. doi:10.4000/cybergeo.3275

Thomas, I., P. Frankhauser, and C. Biernacki. 2008. "The Morphology of Built-Up Landscapes in Wallonia (Belgium): A Classification Using Fractal Indices." *Landscape and Urban Planning* 84(2): 99-115. doi:10.1016/j.landurbplan.2007.07.002

Velázquez, E., I. Martínez, S. Getzin, K. A. Moloney, and T. Wiegand. 2016. "An Evaluation of the State of Spatial Point Pattern Analysis in Ecology." *Ecography* 39(11): 1042-1055. doi:10.1111/ecog.01579

Venables, W. N. and B. D. Ripley. 2002. "Tree-Based Methods." Pp. 251-269 in *Modern Applied Statistics with S*, edited by W. N. Venables and B. D. Ripley. New York, NY: Springer. doi:10.1007/978-0-387-21706-2_9

Walsh, J., J. Watterson, and G. Yielding. 1991. "The Importance of Small-Scale Faulting in Regional Extension." *Nature* 351(6325): 391-393. doi:10.1038/351391a0

White, R. and G. Engelen. 1993. "Cellular Automata and Fractal Urban Form: A Cellular Modelling Approach to the Evolution of Urban Land-Use Patterns." *Environment and Planning A: Economy and Space* 25(8): 1175-1199. doi:10.1068/a251175

White, R., G. Engelen, and I. Uljee. 2015. *Modeling Cities and Regions as Complex Systems: From Theory to Planning Applications*. Cambridge, MA: MIT Press. doi:10.7551/mitpress/9780262029568.001.0001

Wiegand, T. and K. A. Moloney. 2004. "Rings, Circles, and Null-Models

for Point Pattern Analysis in Ecology." *Oikos* 104(2): 209-229. doi:10.1111/j.0030-1299.2004.12497.x

Wu, J., X. Jin, S. Mi, and J. Tang. 2020. "An Effective Method to Compute the Box-Counting Dimension Based on the Mathematical Definition and Intervals." *Results in Engineering* 6: 100106. doi:10.1016/j.rineng.2020.100106

Appendix 1. Derivative of Equation (2)

$$\frac{L}{2^{scale}} = l_{scale}$$

$$2^{-scale} = \frac{l_{scale}}{L}$$

$$\log_2 2^{-scale} = \log_2 \frac{l_{scale}}{L}$$

$$-scale = \log_2 \frac{l_{scale}}{L}$$

$$scale = -\log_2 R_{scale}$$

Appendix 2. Derivative of Equation (3)

$$R_{scale} = \frac{l_{scale}}{L}$$

$$l_{scale} = R_{scale} \times L$$

$$\log_2 OB_{scale} = -FD \times \log_2 l_{scale} + c$$

$$\log_2 OB_{scale} = -FD \times \log_2 (R_{scale} \times L) + c$$

$$\log_2 OB_{scale} = -FD \times \log_2 (R_{scale}) - FD \times \log_2 (L) + c$$

$$\log_2 OB_{scale} = FD \times [-\log_2 (R_{scale})] + c'$$

$$\log_2 OB_{scale} = FD \times scale$$

$$FD = \frac{\log_2 OB_{scale}}{scale}$$

Special Issue

人口學刊
第 65 期，2022 年 12 月，頁 1-41
Journal of Population Studies
No. 65, December 2022, pp. 1-41

分析點分布的臨界尺度：以高雄市登革熱病例的點資料爲例

陳威全*

摘要

疾病的空間群聚分析中常用點資料型態對病例分布位置進行標記與分析。因為點資料的無尺度特性與可調整面單元問題，計算出適當的分析尺度對進行點分布的群聚分析非常關鍵。本研究提出一個資料導向的新分析架構以計算出臨界尺度。此架構是基於兩項經典分析概念：一、點區塊四分樹空間索引方法，二、碎型型態分析中的計算格子法。這兩者都具有捕捉空間跨尺度過程的特性，而這跨尺度特性即為此分析架構的核心概念。本研究應用臺灣高雄市過去近20年的登革熱病例資料作為分析對象，應用前述分析架構計算出每一波疫情年的臨界尺度。透過兩種群聚分析方法，包括核密度推估法與基於密度群聚與噪點分析法，本研究對臨界尺度所計算出來的群聚結構與其他尺度計算所得之群聚結構進行比較與討論。這兩項群聚分析都需設定一距離參數，即其搜尋半徑。因此，本研究透過設定不同的搜尋半徑，借用兩項群聚分析方法所得的群聚型態作為比較不同尺度下的點分布

* 新加坡國立大學地理系研究員
E-mail: wcchin@nus.edu.sg

收稿日期：2022.11.15；接受刊登：2023.03.30

型態以進行比較討論。兩項群聚分析的比較結果都顯示在臨界尺度所計算的群聚結構特性相對較可突出其分布型態。

關鍵詞：空間跨尺度、碎型維度、登革熱、群聚現象、臨界尺度