Geographic Concentration of Industries in Jiangsu, China: A Spatial Point Pattern Analysis using Micro-geographic Data

Xiaoxiang Zhang[a,b], Jing Yao[b]*, Katarzyna Sila-Nowicka[b,c,d], Chonghui Song[e]

[a] *Department of Geographic Information Science, College of Hydrology and Water Resources, Hohai University, Nanjing, China;*

[b] *Urban Big Data Centre, School of Social and Political Sciences, University of Glasgow, Glasgow, UK*

[c] *School of Environment, University of Auckland, Auckland, New Zealand*

[d] *Wroclaw University of Environmental and Life Sciences, Wroclaw, Poland*

[e] *Department of Natural Resources of Jiangsu, Nanjing, China*

*Corresponding author: Jing Yao, Urban Big Data Centre, School of Social and Political Sciences, 7 Lilybank Gardens, University of Glasgow, Glasgow, UK G12 8RZ. Jing.Yao@glasgow.ac.uk

Acknowledgments

This study was supported by the Fundamental Research Funds for the Central Universities (Grant No. B200204029). We are grateful to the Department of Natural Resources of Jiangsu for their provision of the industry land use survey data.

Geographic Concentration of Industries in Jiangsu, China: A Spatial Point Pattern Analysis using Micro-geographic Data

Abstract: Detection of geographic concentration of economic activities at different spatial scales has long been of interest to researchers from spatial economics, regional science and economic geography. Using a unique dataset from the first industrial land use survey of its kind in China, this research is the first effort attempting to explore spatial distribution particularly geographic concentration of industries in China using firm-level data. Distance-based functions and spatial cluster analysis are employed to detect the spatial scales as well as the geographic locations of industrial concentration. The results indicate that four of the five selected industries are in general concentrated in southern Jiangsu at small spatial scales (less than 5 *km*), while the chemical industry demonstrates an overall spatial dispersion pattern relative to the distribution of all other industries. Most industrial clusters have a radius of less than 2.5 *km* containing 20%-60% of enterprises and 60%-86% of employees from each selected industry, with larger clusters showing relatively weaker concentration. This research demonstrates the connections and complementarity of different approaches, complementing previous studies that use distance-based functions with spatial scan statistics.

Keywords: Geographic concentration; Industry agglomeration; Spatial point pattern; Distance-based function; Spatial scan statistics

1 Introduction

It has been well recognized that economic activities tend to locate in certain places (e.g. near market or raw materials) and some industries often cluster or concentrate in certain regions (e.g. technology hubs like Zhongguancun in Beijing, China and Silicon Valley in San Francisco, USA) (Fujita et al., 1999; Combes et al., 2008). The heterogeneous distribution, particularly the tendency of geographic concentration, of economic activities can be attributed to many factors, such as transport costs, labour market pooling, economies of scale, positive externalities and intellectual spillovers (Marshall, 1920; Krugman, 1991). In order to understand various forces shaping the spatial layout of economic activities as well as its implications to economic development and regional inequality, it is often important and necessary to describe and identify the spatial patterns of those activities first, such as geographic concentration or dispersion.

A number of approaches have been proposed to characterize the patterns of economic activities including industrial concentration in geographic space (Duranton and Overman, 2005; Marcon and Puech, 2017), often relying on either areal or point data. Measures based on areal data, such as location quotient (Hoover, 1936), the locational Gini coefficient (Krugman, 1991) and the Ellison–Glaeser index (Ellison and Glaeser, 1997), are often criticized for their sensitivity to the underlying spatial scales and insensitivity to the spatial configuration of territorial units. Although some studies attempted to assess industry agglomeration with spatial statistics such as Moran's I and local indicators of spatial association (LISA) (e.g. Arbia, 2001; Guillain and Le Gallo, 2010), such measures defined on discrete space fail to capture the dispersion or agglomeration of economies across territorial boundaries. In contrast, distance-based methods, such as the K function (Ripley, 1976), the M function (Marcon and Puech, 2010) and the K density (K_d) function (Duranton and Overman, 2005), utilize point data and examine the spatial distribution of economic activities at all spatial scales. However, they cannot identify the exact geographic locations where a pattern (e.g. concentration) occurs.

Existing research on China's industrial distribution primarily relied upon lattice data consisting of administrative districts (e.g. provinces, cities or counties), using approaches such as Gini coefficient (He et al., 2008), the Ellison–Glaeser index (Lu and Tao, 2009), location quotient (Wang et al., 2010), and spatial association indicators such as Moran's I and LISA (Jing and Cai, 2010). While having shed some light on the spatial organization of various industries at certain spatial scales, those methods are inevitably subject to the modifiable areal unit problem (MAUP) (Openshaw, 1984) due to the discrete space under concern.

This research will complement the above work by exploring China's industrial distribution in continuous space, taking Jiangsu – a province on China's east coast – as an example. Jiangsu has long played a leading role in national industrial development. The gross output by industry in Jiangsu was ranked first in China for seven consecutive years up until 2016 (Bureau of Statistics of Jiangsu (BSJ), 2017). However, there has since been intensifying conflict between the limited industrial land and the increasing demand for land resources due to rapid population growth and economic development. In 2016, industrial land accounted for 24.7% of total urban construction land within the province, very close to the highest proportion (25.0%) set by national regulations (BSJ, 2017). In order to meet the industry's sustainable development targets as well as for the overall economy, industrial land utilization in Jiangsu is currently shifting from incremental expansion to land saving and intensive use through industrial restructuring and upgrading (Xinhua News Agency, 2018). Effectively, controlling industrial land expansion and coordinating land resource allocation require a good understanding of the existing spatial distribution of industries within the province.

The aim of this paper therefore is, using a unique firm-level dataset from the first industrial land use survey of its kind in China, to exploit the spatial patterns of industrial distribution within Jiangsu. In addition to distance-based approaches, a spatial scan statistic will be introduced in order to identify the locations of industrial clusters, a measure widely adopted in detecting disease clusters (Kulldorff, 1997). Particularly, this research is intended to investigate (1) at what spatial scales there is spatial concentration of industries, (2) how the degree of concentration varies with spatial scales and (3) where the industrial clusters are if they exist. While the first two questions can be answered by distance-based measures in economic geography, the third question will need to be answered using the spatial scan statistic from epidemiology. Therefore, the primary contribution of this research is the combination of interdisciplinary methods to explore geographic concentration of industries from different perspectives, attempting to identify the linkage and complementarity of different approaches.

The remainder of the paper is organized as follows. The next section briefly reviews the theories related to industrial agglomeration as well as common distance-based approaches for measuring industrial concentration in space. Section 3 describes the study area and industrial data utilized in this research, followed by an introduction of the research methods. Then, the identified spatial patterns of industrial distribution are presented in section 4. Section 5 discusses the complementarity of different approaches and potential forces shaping the identified industrial clusters. The paper concludes with major research findings, highlighting the contribution of this work and the significance of understanding the industrial concentration in Jiangsu for its future economic development.

2 Literature Review

2.1 Theoretical foundations of industrial concentration and empirical findings in China

Spatial concentration of industries can be observed in almost any country or region across the world and has been a classic topic in economics and economic geography since Marshall's study of industrial districts in Britain in the late 19th century (Marshall, 1920). The following review will focus on the theories that are most relevant to this research – neo-classical theory (NCT), new economic geography (NEG) and evolutionary economic geography (EEG). In addition, it should be noted that the central and local governments in China have played a primary role in shaping industrial agglomeration through a range of policies (e.g. tax incentives, subsidized utility and lower land rent) since the economic reforms in 1980s (Fan and Scott, 2003; He et al., 2008; Wei, 2015).

NCT considers that the spread or concentration of economic activities is determined by spatial distribution of exogenous resources such as natural endowments and technologies. Thus, firms tend to concentrate at places with comparative advantages. In China, the transition from centrally-planned to market-driven economy since 1980s has introduced competition among industrial enterprises, encouraging firms to maximize comparative advantages (Fan and Scott, 2003; He et al., 2007). For example, labour-intensive industries such as textiles tended to be spatially concentrated (Lin et al., 2011). Given the uneven distribution of natural resources within Chinese provinces which generally have relatively large geographic area, He et al. (2007) pointed out that resource-based industries might be dispersed across a province but cluster at the city or county level.

NEG focuses on geographic space and the external benefits (externalities), suggesting that location is entirely endogenous and regional agglomeration is a self-reinforcing process

(Krugman, 1991; Fujita et al., 1999). In NEG, agglomeration economies arise because of geographic proximity of firms so that co-located firms can benefit from a pool of skilled labour, supplier-buyer linkages, knowledge spillovers and common resources/infrastructure (e.g. transport, training, and utility supply) (Marshall, 1920; Glaeser et al., 1992). Following the concept of industrial district (Becattini, 1990) and industrial complex (Isard, 1959), Porter (1990) further defined an industrial cluster as a group of related firms and institutions that are geographically concentrated. Instead of comparative advantages, Porter (1998) argued that competitive advantages are more important in the global economy, and clusters can promote productivity and competitive capability through human capital, social networks and face-to-face contact, for which geographic proximity is crucial.

It has been commonly acknowledged that agglomeration economies have major contributions to the development of industrial clusters in China (Fan and Scott, 2003; He et al., 2007; Lu and Tao, 2009). For example, export-oriented firms with high-level foreign investment are largely clustered in the coastal areas with high-quality transportation facilities and network (Fan and Scott, 2003). Strong production linkages were found among the firms of electronic industrial clusters in southern Jiangsu (Wang, 2001) and the information and communication technology (ICT) industrial clusters in Shenzhen (Wang et al., 2010). In addition, various development zones, which were created particularly in order to attract foreign direct investments and boost economy growth, can be seen as a special form of industrial clusters (Hayter, 1997). During 1980s-2000s, the clusters of globalized industries started emerging in the coastal areas (He et al., 2008; Lu and Tao, 2009). Yuan et al. (2017) found that the manufacturing in Nanjing were largely concentrated in the development zones and industrial parks in the suburb.

More recently, EEG suggests that spatial clustering of industries can be explained by genealogy of entrepreneurs (Boschma and Frenken, 2011). Industrial clusters can emerge through the evolutionary process of spinoff formation in the absence of localization economies (Klepper, 2007). Thus, EEG is complementary to the other theories by highlighting the "historical" nature of industrial clusters. For instance, two provinces on the east coast – Jiangsu and Zhejiang – are well-known for their long history of textile production, which has fostered local textile industrial clusters (Wei et al., 2009; Liu, 2014).

Finally, the spatial pattern of concentration usually varies across industries, depending on its type and state intervention (Fan and Scott, 2003; Henderson, 2003). For instance, high-tech innovation clusters tend to be located in large metropolitan areas with little impact of industrialization (Henderson, 2003; Coe et al., 2013). At the national scale, the electronics and telecommunications industry in China are primarily clustered in large metropolitan regions such as Beijing, the lower Yangtze River Delta (YRD) and the Pearl River Delta (PRD) which enjoy skilled labour and capital endowment; and the textile industry is mainly concentrated on the east coast, largely attributed to comparative advantages and agglomeration economies (He et al., 2007; Liu, 2014; Brakman et al., 2017). An example of state intervention is that, in order to better access foreign investment and international markets, the Chinese government encouraged export-intensive industries to cluster on the coastal areas by designating special economic zones with various preferential policies during early stages of the economic reforms in 1980s (Fan and Scott, 2003; Lu and Tao, 2009).

2.2 Distance-based methods applied in industrial distribution studies

In general, distance-based methods compare the spatial distribution of a set of points (e.g. firms) with a reference distribution employed by the null hypothesis assuming no spatial patterns (Diggle, 2003). Such approaches can be grouped into three categories depending on the reference distribution: topological, absolute and relative functions (Marcon and Puech,

2017) (see Table 1).

<Table 1 about here>

Topological functions take geographic space as the reference and the spatial distribution of an economic activity under investigation is compared with complete spatial randomness (CSR). A classic method of this kind is Ripley's *K* function (Ripley, 1976). As the spatial distribution of industries is essentially inhomogeneous (e.g. separated by roads, rivers or mountains), CSR is generally not an appropriate reference. Instead, the difference between two *K* functions, the *D* function (Diggle and Chetwynd, 1991), is often used to represent relative spatial distribution with respect to an alternative (reference) distribution. Some extensions of the *K* function include the K_{mm} function (Penttinen et al., 1992) allowing point weights (e.g. number of employees) and the spatiotemporal *K* function (Arbia, et al., 2010) accounting for the temporal dimension of industrial distribution.

Absolute functions do not consider any references. A typical example is the *K*-density function by Duranton and Overman (2005), K_d, which calculates the density probability of finding a firm at a given distance to the firm of interest. The result of K_d cannot be directly interpreted and is often compared with another group of points using Monte-Carlo simulation. Its variant, K^{emp}, incorporates the number of employees of each firm therefore allowing points (firms) to be weighted.

Relative functions take the distribution of another group of points as the reference. Common approaches include *M*, *m* and *W* functions. The *M* function by Marcon and Puech (2010) counts the neighbours of a targeted industry within a certain distance, comparing its spatial structure with that of all industries, which is the generalization of the *K* function in inhomogeneous space. Using a similar formulation to *M*, the *m* function by Lang et al. (2014)

calculates the ratio of two K_d functions and thus is a density counterpart of M. The W function proposed by Kukuliač and Horák (2017) is defined as the difference of two K_d functions.

It is worth noting that all methods discussed above should be considered complementary rather than substitutional to each other, as they explore different aspects of economic activity distribution over space (Marcon et al., 2015; Marcon and Puech, 2017). Duranton and Overman (2005) suggested that a good index of spatial concentration should meet five criteria: comparable across (1) industries and (2) spatial scales, (3) unbiased, (4) with a reference distribution and (5) allowing statistical significance testing. The measures such as K_d, M, m and W all meet those requirements.

However, one limitation of distance-based methods is that they cannot identify geographic locations of industrial concentration or clusters. That is, they are useful in detecting general spatial patterns of industrial distribution (e.g. concentration or dispersion), but they are less helpful if we would like to know where the industrial agglomeration is if it exists. Therefore, this research intends to complement distance-based functions with spatial scan statistics to answer the three research questions.

3 Materials and Methods

3.1 Study area and data

The study area is Jiangsu province, which is located on the central-east coast of China. As the forth smallest province with an area of 102,600 km^2, Jiangsu however has the highest population density among the 23 Chinese provinces (excluding municipalities) with a total population over 79.9 million (National Bureau of Statistics of China (NBSC), 2017). Furthermore, the gross domestic product (GDP) of Jiangsu has been the second-highest over the last four decades (BSJ, 2017).

The data used in this research come from the first detailed survey of industrial land and enterprises in Jiangsu, which was carried out between June 2016 and June 2017, including the information of nearly 223,000 industrial enterprises such as geographic location, industrial sector and number of employees. Figure 1 shows the number of enterprises at the county level. About 63.2% of industrial enterprises in Jiangsu are located in the three prefectures in the south – Suzhou, Wuxi and Changzhou, which account for 56.9% of overall industrial employees. In contrast, the total proportions of industrial enterprises and employees of the three cities in the north – Huai'an, Suqian and Lianyungang, are only 4.8% and 9.6%, respectively, indicating regional inequalities in industrial development.

<Figure 1 about here>

All enterprises were coded using the industry classification defined by the National Bureau of Statistics (NBSC, 2011) and in total 41 industries were included in the survey. Considering the main characteristics (see Table 2), five manufacturing industries – textile, chemical raw materials and chemical products (abbreviated as chemical), general-purpose machinery (GPM), special-purpose machinery (SPM), and computers, communication equipment and other electronic equipment (CCE) – were selected for subsequent analyses. Each of the five industries was among the top 5 with respect to at least two attributes considered here. Occupying about one third of the province's industrial land, the selected five industries together accounted for 42.6%, 51.8% and 40.5% of all industrial enterprises, employees and gross revenue within the province, respectively.

<Table 2 about here>

3.2 Exploring spatial patterns of industrial distribution

Two relative functions, *M* and *m*, and the spatial scan statistics (Kulldorff, 1997) were adopted in this research. The former two were used to explore the spatial scale of industrial concentration as well as its variation over space (i.e. the first two research questions), and the last one was employed to identify the locations of industrial clusters (i.e. the third research question). *M* and *m* were chosen here because relative measures are often preferred from the economic perspective (Duranton and Overman, 2005; Combes et al., 2008; Ellison et al., 2010). Spatial scan statistics are common approaches for detecting local clusters of spatial events or objects (e.g. disease outbreaks), which were originally proposed for disease surveillance but later have been extensively applied in ecology, demography, psychology and forestry, among others (Kulldorff, 1997, 1999). However, the application of such methods in spatial economics remains very limited.

The *M* function counts the neighbours of an enterprise up to a certain distance and compares with all enterprises within the same distance. Considering the following notation:

i, j: index of enterprises;

N, N_S: the number of all enterprises and the enterprises in industry *S*, respectively;

d_{ij}: distance between i and j;

r: a distance parameter under concern;

z_{ij}: 1 if $d_{ij} \leq r (i \neq j)$; 0 otherwise;

w_i: weight associated with enterprise i;

W, W_S: weight associated with all enterprises and industry *S*, respectively;

The *M* function for industry S can be defined as in (1) (Marcon and Puech, 2010):

$$M_S(r) = \sum_{i=1}^{N_S} \frac{\sum_{j=1}^{N_S} z_{ij} w_j}{\sum_{j=1}^{N} z_{ij} w_j} \Big/ \sum_{i=1}^{N_S} \frac{W_S - w_i}{W - w_i} \qquad (1)$$

where the numerator of formula (1) is the local ratio representing the relative weight (e.g. the number of enterprises or employees) of industry *S* with respect to all industries within a circular area defined by *r*, and the denominator is the global ratio calculated over the entire study area. Therefore, the benchmark of *M* is 1 when the spatial distribution of industry *S* is the same as that of all industries. Accordingly, $M_S(r) > 1$ indicates relative spatial concentration and $M_S(r) < 1$ suggests relative spatial dispersion of *S* at distance *r*. In this research, the weight was defined as the number of employees for each enterprise and industry.

The *m* function is defined in the same way as *M*. The only difference is that the z_{ij} in (1) is replaced by a kernel function, k_{ij}, which estimates the number of neighbours of the ith enterprise at distance *r*. Following Duranton and Overman (2005) and Lang et al. (2014), a Gaussian kernel was employed here with the optimal bandwidth obtained by the approach from Silverman (1986). The *m* function also has a benchmark value 1 and the interpretation of *m* values is similar to that of *M*, except that *m* counts the neighbours at, rather than up to, a distance, which makes *m* a density rather than cumulative function like *M*.

In this research, both *M* and *m* functions were calculated up to 60 *km* – the distance of half an hour's drive at the national speed limit (120 *km*/hour) on China's highways, which was considered by local authorities as the maximum radius defining the spatial extent applicable to local industrial concentration within Jiangsu. The statistical significance of *M* and *m* values were tested against a series of random distributions generated by Monte–Carlo simulations, where enterprises were redistributed across actual geographic locations. In this research, global confidence intervals at 5% significance level with 1,000 simulations were derived using the commonly used procedure by Duranton and Overman (2005) (see Marcon and Puech (2010) and Lang et al. (2014)).

The particular spatial scan statistical approach employed here is the Bernoulli probability model designed for data consisting of two groups: cases and controls, where a circular window with varying sizes scans over the study area to detect regions with significantly high or low proportion of cases (Kulldorff, 1997). Although it is possible to use other shapes of scanning window, circles are adopted here to enable the comparison with the results from distance-based functions which count neighbours in a similar way. Specifically, cases represent the employees from the industry of interest and controls refer to the employees from all the other forty industries included in the survey (i.e. the rest of the economy). The null hypothesis is that the probability of finding a case inside and outside the window is same. The test statistic for a particular window can be expressed as in (2) (Kulldorff, 1997):

$$\lambda = a\left(\frac{n_S}{n}\right)^{n_S}\left(1-\frac{n_S}{n}\right)^{n-n_S}\left(\frac{N_S-n_S}{N-n}\right)^{N_S-n_S}\left(1-\frac{N_S-n_S}{N-n}\right)^{(N-n)-(N_S-n_S)} \tag{2}$$

Where n and n_S are the number of all enterprises and the enterprises in industry S within the scanning window, respectively; N and N_S are defined as before; the value of a is 1 if $\frac{n_S}{n} > \frac{N_S-n_S}{N-n}$ when seeking clusters with high ratio of cases (i.e. geographic concentration of the industry under concern). The statistical significance of λ is tested by Monte Carlo simulation. In this research, five separate analyses were implemented, each dichotomizing the data into one of the five interested industries and the other industries.

The above approaches were implemented with several software tools. The *M* and *m* functions were carried out using the *dbmss* package in *R* (Marcon et al., 2015). The spatial scan statistics were calculated in SaTScan (https://www.satscan.org/). The commercial GIS software ArcGIS 10.6 (Environmental Systems Research Institute, Redlands, California, USA) was utilized for spatial data processing, management and visualization.

4 Results

4.1 Spatial scales of industrial concentration

The values of the *M* and *m* functions are described by Figures 2 and 3, respectively. In each graph the solid line indicates the *M* or *m* values calculated using the actual enterprise locations, the red broken line represents the expected *M* or *m* values as if the industry under concern has the same spatial distribution as that of all industries, and the grey area is the confidence envelope at the 5% significance level.

<Figure 2 about here>

All industries except the chemical industry demonstrate significant geographic concentration patterns over all the underlying distances, although the degree of which varies with the distance and across industries (see Figure 2). For the textile industry (Figure 2(a)), the higher *M* values (>4) are obtained at shorter distances (<650 *m*). For example, a *M* of value 4.1 implies that the relative density of neighbours (employees) from the textile industry within 500 *m* is about 4.1 times higher than it would be if the textile enterprises were distributed as they are in the whole industry. Figure 2(a) shows that the *M* value decreases from 3 to 2 when the size of neighbourhood increases from 5 *km* to 27 *km*. Similarly, the SPM industry also has higher concentration (*M* > 4) at shorter distances (< 450 *m*), but the *M* values drops more quickly below 2, at a distance of 6 *km* (Figure 2(d)). Comparatively, all the *M* values of the GPM and the CEE industries are lower than 3 and 2 (Figures 2(c) and 2(e)), respectively, indicating weaker industrial concentration. But the overall monotonically decreasing pattern of the former is similar to those of the textile and the SPM industries, and the latter has two transition points – one at 2 *km* with the *M* value of 1.38 and the other at 21 *km* with the *M* value of 1.62. Compared with the other four industries, the chemical industry seems having very different spatial distribution as the geographic concentration disappears when the distance is larger than 12 *km* and all the *M* values less than 1 are statistically significant, indicating that chemical enterprises are dispersed over space relative to the distribution of all other industries. Also, Figure 2(b) shows the strongest concentration of chemical enterprises (*M* is 1.86) occurs at a distance of 1 *km*.

<Figure 3 about here>

Figure 3 suggests that the *m* values for all industries are less than the corresponding *M* values at the same distance. Again, all industries except the chemical industry have prominent spatial concentration particularly at smaller distances, which however starts disappearing from a distance between 40 *km* and 60 *km*. Among all industries, the textile industry has the highest local density of same-type neighbours at all distances within 20 *km* (Figure 3(a)), indicating the existence of textile enterprise (employee) concentration at spatial scales (radiuses) smaller than 20 *km*. For instance, the value of *m*, 2.13, at the distance of 10 *km* implies that the proportion of the neighbouring employees from the same industry at this distance is 113% higher than that in the entire province. Although weaker spatial concentration still exists at larger scales (i.e. distance > 20 *km*), it is not statistically significant when the distance is beyond 47 *km*. The GPM and CCE industries seem have similar concentration patterns (Figures 3(c) and 3(e)): the *m* values are largely between 1.5 and 1.75 up to the distance of 20 *km*, which decline faster for the former at distances exceeding 20 *km* but are not significant anymore beyond 52 *km*; for the latter, the *m* values are not significant at the distances of 54-58 *km*, but show dispersion pattern beyond that. Compared with textile, GPM and CCE industries, the SPM industry (Figure 3(d)) shows the weakest concentration at distances up to 48 *km*, with all *m* values less than 1.5. Similar to the *M* values in Figure 2(b), the *m* values in Figure 3(b) indicate a very different spatial distribution of the chemical industry from the other four industries, which only has subtle concentration at very small scales (distances less than 2 *km*) and significant dispersion pattern beyond the distance 6 *km*.

4.2 Spatial clusters of industries

First, the number of local clusters varies across industries. Table 3 indicates that the GPM and the SPM industries have more clusters (185 for each) than the others: 43 (textile), 63 (chemical) and 50 (CCE). Also, all industries except chemical have more than 40.0% of enterprises located in the clusters. Although the clusters of chemical industry contain only 19.7% of relevant enterprises, it includes over 80.0% of all employees from that industry. The CCE industry has the highest proportion (85.9%) of employees inside its clusters, the value of which is between 60% and 70% for the other three industries (i.e., textile, GPM and SPM).

<Table 3 about here>

Obviously, those clusters are not evenly distributed over space as shown in Figures 4 and 5, most of which are in the south, particularly within Suzhou, Wuxi and Changzhou (Figure 4). Comparatively, the northern Jiangsu has less clusters which are more sparsely distributed. As for each industry, Figure 5(a) indicates that the smaller clusters of the textile industry are largely in the south while the larger ones are in the north and along the east coast. For the chemical industry, most smaller clusters are scattered in the southern cities and there is a group of clusters in the northern Nanjing (Figure 5(b)). The GPM and the SPM industrial clusters seem having similar spatial distribution, which are largely located in the central and southern Jiangsu with a few in the northwest (Figures 5(c) and 5(d)). In contrast, the CCE industry has two distinct groups of clusters mainly in Suzhou and near the border of Nanjing and Zhenjiang, with few clusters in the northern and central Jiangsu.

<Figure 4 about here>

<Figure 5 about here>

Further, the selected five industries have different degrees of relative concentration across space. As shown in Table 3, the GPM industry has the weakest cluster where the proportion of its employees inside that cluster is about 1.5 times that outside, and the SPM industry has the strongest cluster with a value 15.3. On average, the GPM and the SPM industries have higher degrees of industrial concentration with mean values of 8.1 and 10.7, respectively. Overall, there are less variations of relative concentration in the chemical and the CCE industries with standard deviations of 1.1 and 1.2, respectively. Figure 4 suggests that all the five industries have a small proportion (2%-11%) of the clusters with the lowest level of relative concentration (i.e. values less than 2.50), and only the GPM and the SPM industries have clusters with the highest degree of relative concentration (i.e. values larger than 10). In contrast, the clusters of chemical and CCE industries show weaker relative concentration, with all values less than 7.5.

In terms of spatial variations, Figure 5 shows that most clusters with stronger relative concentration are located in the south. For example, there is a distinct cluster with a high degree of textile enterprise concentration in Suzhou, while most clusters with the weakest concentration of textile enterprises are in the central east (Figure 5(a)). As to the chemical industry, the clusters with higher level of relative concentration are largely along the Yangtze River and the east coast, while those with the least concentration are mainly in the north (Figure 5(b)). Both GPM (Figure 5(c)) and SPM (Figure 5(d)) industries have a range of clusters of varying size and degree of relative concentration in the central and southern province. Figure 5(e) indicates that most clusters of the CCE industry are in Suzhou and Nanjing with relative concentration values between 2.51-7.5.

Finally, both Figures 5 and 6 show the varying sizes of the industrial clusters. It seems in

Figure 5 that larger clusters tend to have weaker relative concentration. Among all industrial clusters, only two have a size larger than 60 *km* – the searching radius used in the *M* and the *m* functions: one in the textile industry (r = 85.2 *km*) (Figure 5(a)) and the other in the GPM industry (r = 60.2 *km*) (Figure 5(c)). Considering the visual effect, Figure 6 depicts the size distribution of the clusters with a radius of less than 10 *km* (around 91.0% of all identified clusters). Most clusters virtually have a radius smaller than 2,500 *m*, the share of which is about 78.0% (textile and chemical), 80.0% (GPM), 84.0% (SPM) and 79.0% (CCE). Also, the median cluster sizes of GPM and SPM industries are very similar and much smaller than those of the other three industries. In addition, the clusters of SPM industry have the lowest mean size (1372.8 *m*), about 386.4 *m* less than the largest value of the CCE industry.

<Figure 6 about here>

5 Discussion

5.1 Complementarity of different approaches

The three approaches adopted in this research explore spatial concentration of industries from different perspectives. As Marcon and Puech (2010) pointed out, *m* as a density function can measure local density of industries more precisely while *M* values can better reflect the spatial structure of industries. As the aggregations at smaller distances can affect the estimates at larger distances due to the nature of cumulative functions, it is not surprising that the *M* values are generally larger than the *m* values (see Figures 2 and 3). Both *M* and *m* functions detected distinct industrial concentration in Jiangsu except for the chemical industry, where the largest spatial scale of concentration identified by the former is larger than that by the latter because *m* focuses on the local scale. Both *M* and *m* values show that the chemical industry is mostly and strongly characterized by spatial dispersion rather than concentration. The results from the spatial scan statistics (see Figure 5) are consistent with those from both *M* and *m* functions in the sense that stronger industrial concentrations mostly occur at smaller spatial scales particularly at distances less than 5 *km*, which is further verified by Figure 6.

The results obtained from different techniques can be linked to each other to better understand industrial distribution over space. For example, both Figures 2(b) and 3(b) suggest that spatial distribution of the chemical industry is very different from those of the other four industries, i.e. with strong overall spatial dispersion pattern. Meanwhile, Figure 5 clearly shows that the CCE industrial clusters have a different spatial distribution from the others. In fact, Table 3 indicates that although the local clusters of both (i.e. chemical and CCE) contain over 80% employees of the respective industry, the share of involved enterprises is 19.7% for the former while 43.4% for the latter. Again, it can be observed that the curve in Figure 3(b) drops quickly below 1.0 and the one in Figure 3(e) declines more slowly until reaches the distance 54 *km* when *m* is no longer significant. This implies that the chemical industry could be concentrated in a few big enterprises with a large number of employees while the CCE industries could be concentrated in many smaller enterprises. For example, Nanjing, where three clusters were identified (see Figure 5(b)), has about 600 chemical enterprises, while Suzhou, where most CCE clusters were found, has over 4,000 relevant enterprises, where both cities have similar number of employees (about 1.4 million) in the respective industry (BSJ, 2017). Therefore, the three methods are complementary to each other and the conjoint use of them can offer a better understanding of industrial distribution.

5.2 Reflections on the driving forces of identified industrial clusters

The following discussion is limited to the theories and the role of Chinese government mentioned before. Specifically, the associated influencing factors include comparative advantages, agglomerative forces, history and the development zone initiative in China.

There is no intention here to provide a comprehensive discussion on all relevant potential driving forces of the identified industrial clusters.

First, the local clusters of textile and chemical industries can be partially attributed to the comparative advantages associated with natural resources and labours. Locating along the Yangtze River and the east coast meets the need of the chemical and textile industries for large water consumption and wastewater treatment. Also, the textile industry has high requirements for climatic conditions and are often located in warm and humid areas such as the lower YRD region. In addition, the textile industry is labour-intensive and cities in southern Jiangsu benefit from a large pool of local skilled labour due to the long history of textile production in those areas. In particular, the three cities – Nantong, Suzhou and Wuxi – together account for 76.6% of all employees in the textile industry within the province.

Second, agglomerative forces such as dedicated infrastructure and intermediate industries play a crucial role in the clustering of the selected industries under concern. Jiangsu have well-developed transportation network including expressways, railways and waterways, greatly facilitating the trade with the rest of the country and the world particularly for the export-intensive textile and chemical industries. For the CCE clusters in Suzhou, it is common to find the co-location of manufacturers with suppliers and assemblers, and thus the firms can benefit from reduced transport and transaction costs and focus on their core competencies through vertical disintegration (Coe et al., 2013).

Third, the evolutionary perspective by EEG can help explain the existence of local clusters of textiles, GPM and SPM industries. Textile is a traditional industry which can be dated back more than two thousand years ago in Jiangsu, and history plays a critical role in shaping the local clusters of textile production in southern Jiangsu, along the Yangtze River and the east coast. The majority of the firms of GPM and SPM industries are developed from the township and village enterprises in southern Jiangsu in 1980s. The process of marketization and globalization has helped produce contemporary distribution of those two industries (Wei, 2010). For example, the four cities in the south – Changzhou, Nantong, Suzhou and Wuxi – account for 71.7% and 63.7% of all employees in the GPM and the SPM industries, respectively.

Finally, development zones have encouraged the formation of local industrial clusters particularly in the chemical and CCE industries. The clusters in northern Nanjing and south-western Yangzhou (see Figure 5(b)) locate two large development zones – Nanjing Chemical Industrial Park and Yangzhou Chemical Industrial Park, and the two cities together have about two thirds of the chemical industry's employees in the province. Benefited from the proximity to Shanghai (about 110 *km*) – the largest and most prosperous city in China, the development zones in Suzhou have attracted a large amount of foreign direct investment particularly on its high-tech industries like CCE (Wei et al., 2009), which has greatly enhanced industrial concentration in Suzhou. In fact, about half of the provincial CCE employees work in development zones and Suzhou has a share of 56.2% of such employees.

Of course, the above reflections need to be formally tested using confirmatory analyses such as spatial regression. This leads to possible future work. First, hypotheses can be formulated based on the identified industrial clusters to examine the potential driving forces, which can be tested against classic theories such as NCT and NEG. Also, the influences of industrial concentration on rural-urban/intra-urban migration, employment opportunities and income variations, etc. can be investigated to inform policy-making towards coordinated regional development. Third, the alternative scanning window (i.e. ellipse) can be used in the spatial scan statistics to explore different shapes of industrial clusters. Finally, the computational efficiency of the techniques adopted here needs improvement*.

There are some limitations of this research. First, current work only includes five selected industries, failing to provide a complete picture of the industrial distribution within Jiangsu. Second, this study only considers co-located firms from the same industry and therefore it is not clear whether the identified industrial clusters benefit from Marshallian specialisation, Jacobian diversification or both. Finally, this research is limited to a single province which fails to capture the industrial agglomeration at larger spatial scales involving the surrounding areas such as Zhejiang Province and Shanghai.

6 Conclusions

Using a unique micro-geographic dataset of industrial enterprises, this research explores the spatial scales and geographic locations of industrial concentration in Jiangsu, China. The results indicate that the scale and degree of industrial concentration vary across space and industries. The five selected industries generally concentrate at small spatial scales (less than 5 *km*) with over 60% (80% for the chemical and the CCE industries) employees inside local industrial clusters. Particularly, the chemical industry mainly concentrates in larger enterprises with more employees and the CCE industry is primarily clustered in smaller enterprises having less employees. Also, the chemical industry demonstrates an overall dispersion pattern across the province. This research demonstrates the connections and complementarity of different approaches, complementing previous studies relying on distance functions with local spatial cluster analysis. As a major member of the YRD Economic Zone, the industries in Jiangsu play a primary role in the economic development of the YRD region.

* Generally, searching local clusters for large datasets with spatial scan statistics can be computationally expensive. For instance, a typical spatial cluster analysis using the SaTScan software involving 250,000 observations requires a computer memory of 128 GB (Kulldorff, 2018). For the dataset used in this research which contains about 223,000 enterprises, it took about 20-80 minutes for running 1000 simulations of the M function, 18-22 hours for the m function, and nearly 4 hours to identify local clusters for each industry using a desktop with the Intel Xeon Processor E5-2640@2.60 GHz and 256 GB memory.

Understanding the scales and locations of industrial concentration remains crucial if the full advantages of economic agglomeration are to be taken.

Declaration of interest statement

No potential conflict of interest was reported by the authors.

References

Arbia, G. (2001). The role of spatial effects in the empirical analysis of regional concentration. *Journal of Geographical Systems*, 3(3), 271-281.

Arbia, G., & Espa, G. (1996). *Statistica Economica Territoriale*. CEDAM, Padova.

Becattini, G. (1990). The Industrial District as a Creative Milieu. In *Industrial Change and Regional Development*, edited by G. Benko and M. Dunford, 102–116. London: Belhaven Press.

Boschma, R., & Frenken, K. (2011). The emerging empirics of evolutionary economic geography. *Journal of Economic Geography*, 11(2), 295-307.

Brakman, S., Garretsen, H., & Zhao, Z. (2017). Spatial concentration of manufacturing firms in China. *Papers in Regional Science*, 96, S179-S205.

Bureau of Statistics of Jiangsu. (2017). *Jiangsu Statistical Yearbook 2017*. China Statistics Press. Beijing, China. (in Chinese).

Coe, N. M., Kelly, P. F., & Yeung, H. W. (2013). *Economic geography: a contemporary introduction (2nd ed.)*. John Wiley & Sons.

Combes, P. P., Mayer, T., & Thisse, J. F. (2008). *Economic geography: The integration of regions and nations*. Princeton University Press.

Diggle, P. J. (2003). *Statistical analysis of spatial point patterns (2nd edition)*. London, UK: Edward Arnold.

Duranton, G., & Overman, H. G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, 72(4), 1077-1106.

Ellison, G., & Glaeser, E. L. (1997). Geographic concentration in US manufacturing industries: a dartboard approach. *Journal of Political Economy*, 105(5), 889-927.

Fan, C. C., & Scott, A. J. (2003). Industrial agglomeration and development: a survey of spatial economic issues in East Asia and a statistical analysis of Chinese regions. *Economic Geography*, 79(3), 295-319.

Fujita, M., Krugman, P. R., & Venables, A. J. (1999). *The spatial economy: Cities, regions, and international trade*. MA: MIT Press.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, 100(6), 1126-1152.

Guillain, R., & Le Gallo, J. (2010). Agglomeration and dispersion of economic activities in and around Paris: an exploratory spatial data analysis. *Environment and Planning B: Planning and Design*, 37(6), 961-981.

He, C., Wei, Y. D., & Pan, F. (2007). Geographical concentration of manufacturing industries in China: The importance of spatial and industrial scales. *Eurasian Geography and Economics*, 48(5), 603-625.

He, C., Wei, Y. D., & Xie, X. (2008). Globalization, institutional change, and industrial location: Economic transition and industrial concentration in China. *Regional Studies*, 42(7), 923-945.

Henderson, J. V. (2003). Marshall's scale economies. *Journal of Urban Economics*, 53(1), 1-28.

Hayter, R. (1997). *The Dynamics of Industrial Location: The Factory, the Firm and the Production System*. Chichester: John Wiley & Sons.

Hoover, E. M. (1936). The measurement of industrial localization. *The Review of Economic Statistics*, 162-171.

Isard, W. (1959). *Industrial Complex Analysis and Regional Development: A Case Study of Refinery-Petrochemical-Synthetic Fiber Complexes and Puerto Rico*. Cambridge, MA: MIT Press.

Jing, N., & Cai, W. (2010). Analysis on the spatial distribution of logistics industry in the developed East Coast Area in China. *The Annals of Regional Science*, 45(2), 331-350.

Kopczewska, K. (2017). Distance-based measurement of agglomeration, concentration and specialisation. In *Measuring Regional Specialisation: A New Approach* (pp. 173-216). Cham, Switzerland: Springer.

Klepper, S. (2007). Disagreements, spinoffs, and the evolution of Detroit as the capital of the US automobile industry. *Management science*, 53(4), 616-631.

Krugman, P. (1991). *Geography and Trade*. Cambridge, MA: MIT Press.

Kukuliač, P., & Horák, J. (2017). W Function: A New Distance-Based Measure of Spatial Distribution of Economic Activities. *Geographical Analysis*, 49(2), 199-214.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics-Theory and methods*, 26(6), 1481-1496.

Kulldorff, M. (1999). Spatial scan statistics: models, calculations, and applications. In *Scan statistics and applications* (pp. 303-322). Birkhäuser, Boston, MA.

Kulldorff, M. (2018). *SaTScan Users Guide for version 9.6*. https://www.satscan.org/cgi-bin/satscan/register.pl/SaTScan_Users_Guide.pdf?todo=process_userguide_download (Accessed on 7 June 2019)

Lang, G., Marcon, E., & Puech, F. (2014). *Distance-Based Measures of Spatial Concentration: Introducing a Relative Density Function*. Working Papers, HAL (version 1). https://hal-mnhn.archives-ouvertes.fr/INRA/hal-01082178v1 (Accessed 29/03/2019)

Liu, Z. (2014). Global and local: Measuring geographical concentration of China's manufacturing industries. *The Professional Geographer*, 66(2), 284-297.

Lin, H. L., Li, H. Y., & Yang, C. H. (2011). Agglomeration and productivity: Firm-level evidence from China's textile industry. *China Economic Review*, 22(3), 313-329.

Lu, J., & Tao, Z. (2009). Trends and determinants of China's industrial agglomeration. *Journal of Urban Economics*, 65(2), 167-180.

Marcon, E., & Puech, F. (2010). Measures of the geographic concentration of industries: improving distance-based methods. *Journal of Economic Geography*, 10(5), 745-762.

Marcon, E., & Puech, F. (2017). A typology of distance-based measures of spatial concentration. *Regional Science and Urban Economics*, 62, 56-67.

Marcon, E., Traissac, S., Puech, F., & Lang, G. (2015). Tools to characterize point patterns: dbmss for R. *Journal of Statistical Software*, 67(3), 1-15.

Marshall, A. (1920). *Principles of Economics*. London: MacMillan.

National Bureau of Statistics of China. (2011). *Industrial classification for national economic activities (GB/4754-2011)*. http://www.stats.gov.cn/tjsj/tjbz/hyflbz/2011/ (Accessed on 7 June 2019). (in Chinese).

National Bureau of Statistics of China. (2017). *China Statistical Yearbook 2017*. China Statistics Press. Beijing, China. (in Chinese).

Openshaw, S. (1984). *The modifiable areal unit problem*. Norwick: Geo Books.

Porter, M. E. (1990). The competitive advantage of nations. New York: Free Press.

Porter, M. E. (1998). Clusters and the new economics of competition. *Harvard Business Review*, 76(6), 77-90.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13(2), 255-266.

Silverman B. W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall, London.

Wang, J. (2001). *Innovative spaces: Enterprise clusters and regional development*. Beijing: Peking University Press. (in Chinese).

Wang, C. C., Lin, G. C., & Li, G. (2010). Industrial clustering and technological innovation in China: new evidence from the ICT industry in Shenzhen. *Environment and Planning A*, 42(8), 1987-2010.

Wei, Y. D. (2010). Beyond new regionalism, beyond global production networks: remaking the Sunan model, China. *Environment and Planning C: Government and Policy*, 28(1), 72-96.

Wei, Y. D. (2015). Zone fever, project fever: Development policy, economic transition, and urban expansion in China. *Geographical Review*, 105(2), 156-177.

Wei, Y. H., Lu, Y., & Chen, W. (2009). Globalizing regional development in Sunan, China: does Suzhou Industrial Park fit a neo-Marshallian district model?. *Regional Studies*, 43(3), 409-427.

Xinhua News Agency. (2018). *Promoting land saving and intensive use in Jiangsu*. http://www.xinhuanet.com/2018-07/06/c_1123089318.htm (Accessed on 20 June 2020). (in Chinese).

Yuan, F., Gao, J., Wang, L., & Cai, Y. (2017). Co-location of manufacturing and producer services in Nanjing, China. *Cities*, 63, 81-91.

Table 1 Common distance-based methods for measuring geographic concentration

Types of Function	Method	Related Literature	Main Feature	Advantages/Disadvantages
Topological	K function	(Ripley, 1976)	With a null hypothesis of CSR	**Advantages** (1) Can detect spatial patterns in continuous space, i.e., at all spatial scales; (2) Provide statistical significance tests of the measures. **Disadvantage** (3) Cannot identify the locations where economic activities concentrate.
	D function	(Diggle and Chetwynd, 1991)	Difference between two K functions; for non-stationary point patterns.	
	K_{mm} function	(Penttinen et al., 1992)	Allow point weights (e.g. number of employees)	
	the spatiotemporal K function	(Arbia, et al., 2010)	Consider the temporal dimension of industrial distribution	
Absolute	K density (K_d) function	(Duranton and Overman, 2005)	A probability density function	
	K^{emp} function	(Duranton and Overman, 2005)	Allow point weights (e.g. number of employees)	
Relative	M function	(Marcon and Puech, 2010)	A cumulative function	
	m function	(Lang et al., 2014)	A density counterpart of M function	
	W function	(Kukuliač and Horák, 2017)	the difference of two K_d functions	

Table 2 Main characteristics of selected industries

Code	Industry	Abbreviation	Land use area		Number of enterprises		Number of employees		Gross revenue	
			km^2	**Rank***	**1000**	**Rank**	**10,000**	**Rank**	**Billion ¥**	**Rank**
17	Textile	Textile	200.1	4	19.5	4	106	2	381.7	7
26	Chemical raw materials and chemical products	Chemical	233.5	2	10.1	9	58	8	757.6	2
34	General-purpose machinery	GPM	256.8	1	32.8	1	92	3	580.7	4
35	Special-purpose machinery	SPM	214.1	3	23.3	2	77	4	365.7	9
39	Computers, communication equipment and other electronic equipment	CCE	135.4	8	9.2	10	158	1	1094.6	1

*: The rank is obtained by comparison across all industries.

Table 3 Information of local clusters from spatial scan statistics

Industry	Number of high-ratio clusters	% of enterprises within clusters	% of employees within clusters	Relative Concentration*			
				min	max	mean	standard deviation
Textile (Textile)	43	56.3	66.5	1.8	9.5	6.8	2.8
Chemical raw materials and chemical products (Chemical)	63	19.7	82.5	1.6	6.4	5.5	1.1
General-purpose machinery (GPM)	185	43.3	66.9	1.5	10.8	8.1	2.9
Special-purpose machinery (SPM)	185	49.5	62.7	2.3	15.3	10.7	4.6
Computers, communication equipment and other electronic equipment (CCE)	50	43.4	85.9	2.0	6.1	4.9	1.2

*: The relative concentration is defined as the ratio between the shares of cases inside and outside a cluster

Figures

Figure 1 Spatial distribution of industry enterprises across prefectures in Jiangsu

Figure 2 Results of *M* functions for the five selected industries: (a) Textile; (b) Chemical; (c) GPM; (d) SPM; (e) CCE

Figure 3 Results of *m* functions for the five selected industries: (a) Textile; (b) Chemical; (c) GPM; (d) SPM; (e) CCE

Figure 4 Cluster centers of five selected industries

Figure 5 Local clusters identified by spatial scan statistics: (a) Textile; (b) Chemical; (c) GPM; (d) SPM; (e) CCE

Figure 6 Size variation of local clusters with a radius of less than 10 *km*

Figure 1 Spatial distribution of industry enterprises across prefectures in Jiangsu

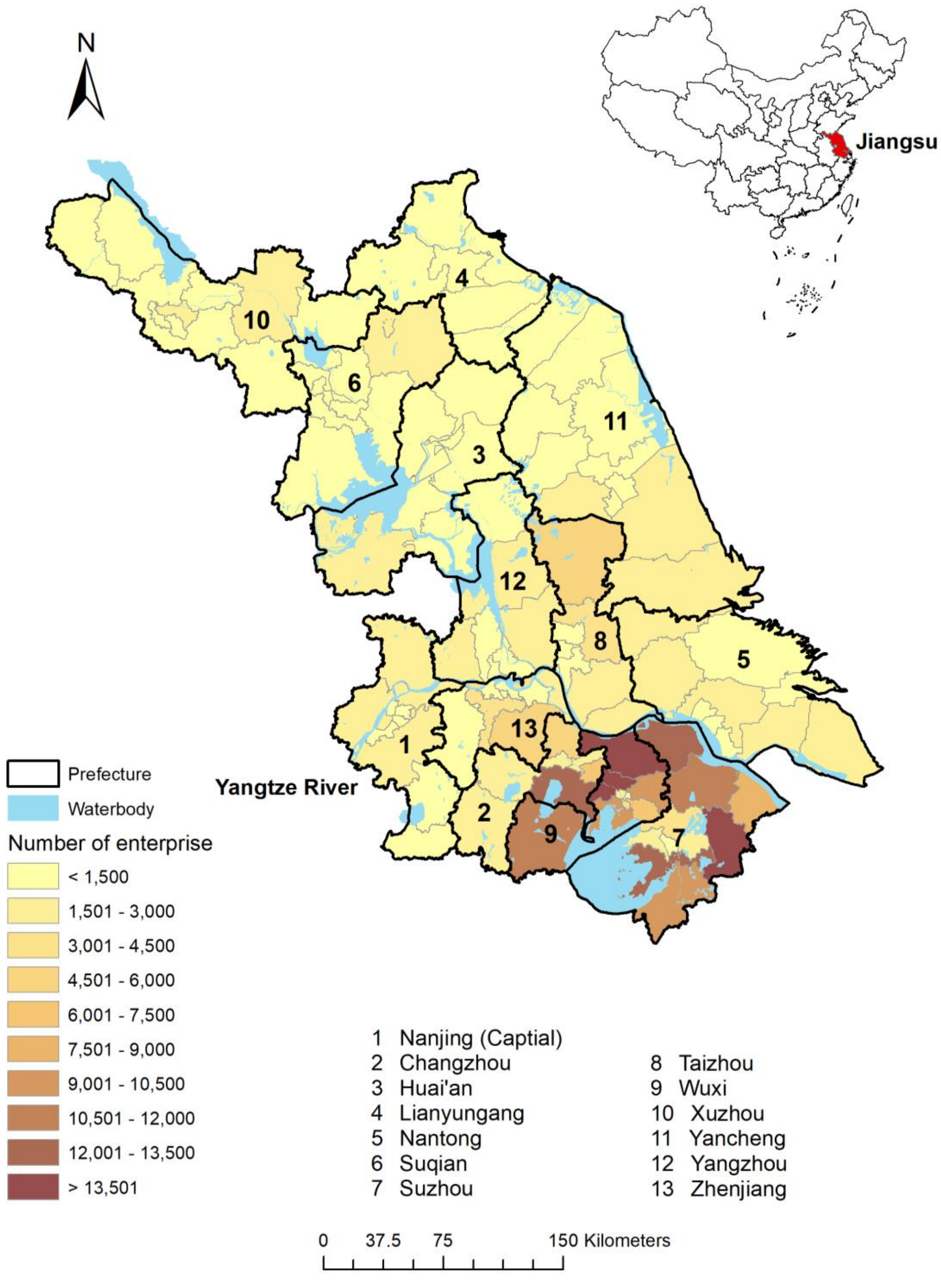


Figure 2 Results of *M* functions for the five selected industries: (a) Textile; (b) Chemical; (c) GPM; (d) SPM; (e) CCE

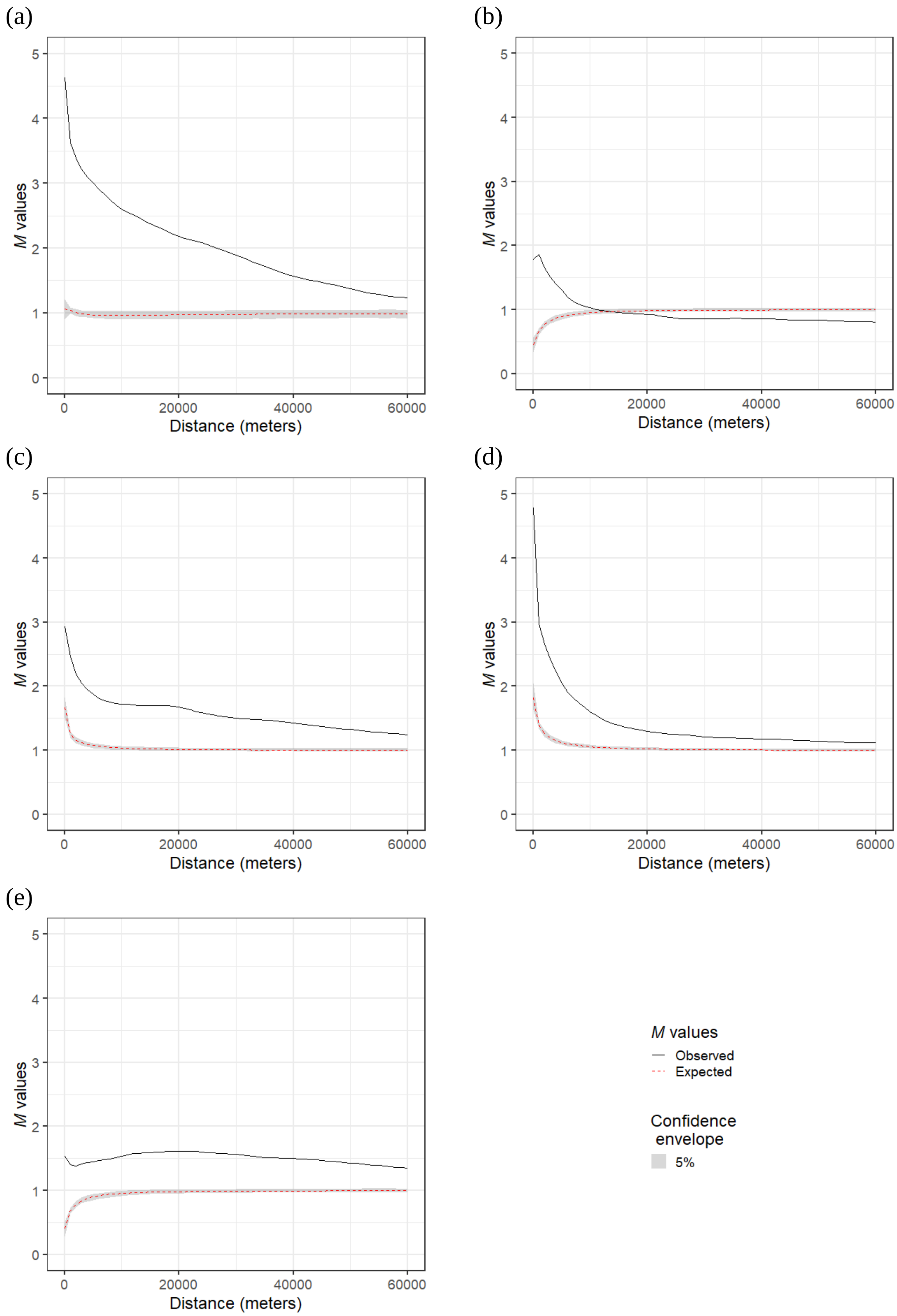


Figure 3 Results of *m* functions for the five selected industries: (a) Textile; (b) Chemical; (c) GPM; (d) SPM; (e) CCE

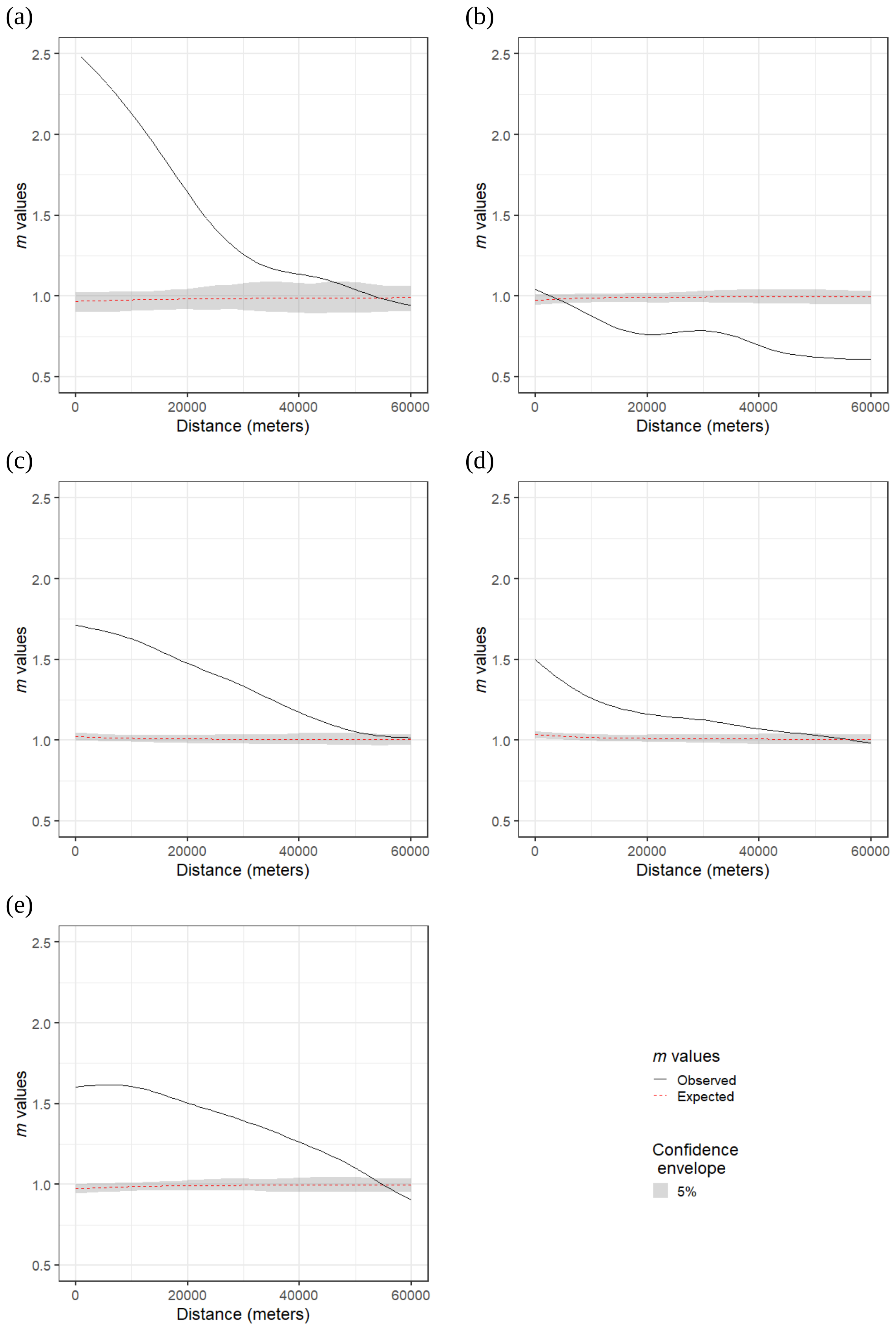
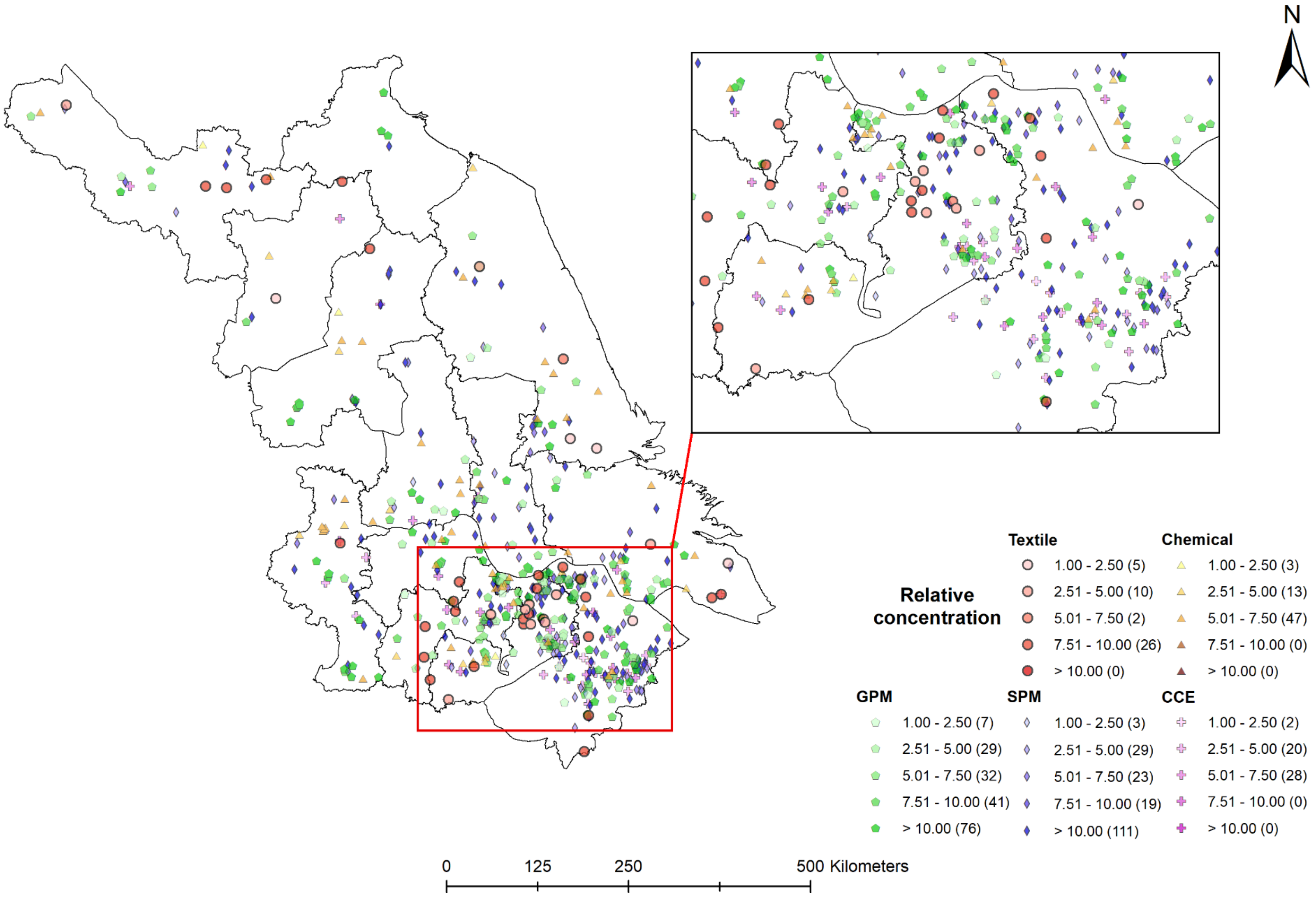


Figure 4 Cluster centers of five selected industries†



† The digits inside the brackets in the legend denote the number of clusters in the corresponding categories

Figure 5 Local clusters identified by spatial scan statistics: (a) Textile; (b) Chemical; (c) GPM; (d) SPM; (e) CCE

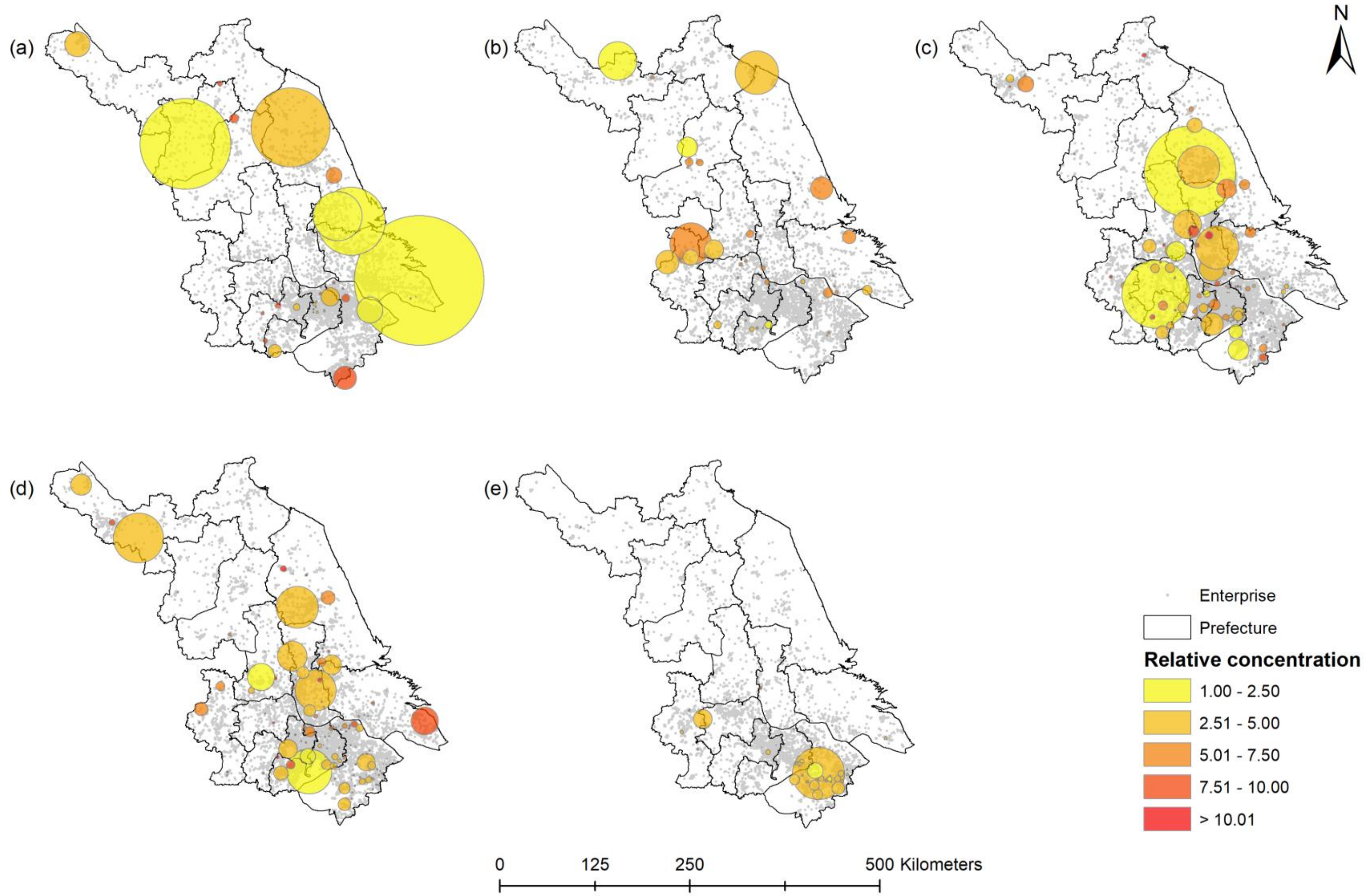


Figure 6 Size variation of local clusters with a radius of less than 10 *km*