IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 10, 2017, accepted May 23, 2017, date of publication May 26, 2017, date of current version June 27, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2708762

Identifying Urban Subcenters from Commuting Fluxes: A Case Study of Wuhan, China

ZHONGLIANG FU[1,2], KAICHUN ZHOU[1], AND LIANG FAN[1]
[1]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[2]Collaborative Innovation Center of Geospatial Technology, Wuhan 430079, China

Corresponding author: Kaichun Zhou (kaichun_zhou@foxmail.com)

ABSTRACT It is of interest to analyze urban spatial structure by identifying urban subcenters, for which published literature proposes many methods. Although these methods are widely applied, they demonstrate obvious shortcomings that restrict further application. Therefore, it is of great value to propose a new urban subcenter identification method that can overcome these shortcomings. In this paper, we introduce an alternative method. Unlike two-stage procedures and other arbitrary methods, our method is not based on arbitrary cutoff values and is entirely parameter free. We first calculate the commuting fluxes for each pair of census tracts and use the fluxes to represent a local density. After that, the census tracts are partitioned into several clusters using a clustering algorithm. Finally, subcenters are derived from the clusters through a circularly shaped spatial scan statistic. We apply this method to 2010 and 2015 census data sets for Wuhan, China. The identification and comparison results demonstrate that our proposed method is effective and can be applied toward future research.

INDEX TERMS Urban subcenter, commuting flux, cluster, spatial scan statistic.

I. INTRODUCTION

Urban spatial structure has been studied extensively in order to determine the 'qualitative change' in urban patterns and spatial structures [1], [2]. Concordant with the development of a city, the form of the spatial structure often evolves from monocentric to polycentric [3]–[5].

A monocentric city possesses a single center. All of the activities are concentrated within the central business district (CBD), and the employment (or population) density usually decentralizes away from it. In contrast, a polycentric city usually has one or more subcenters in addition to the CBD [6].

Rapid urbanization has two byproducts: high productivity and overcrowding [7]. To maintain sustainable development under an increasing population pressure [8], it is of substantial importance to study urban spatial structure. Accordingly, since the urban subcenter represents the concentration area of the population and resources, studying the urban subcenter is an effective way to analyze urban spatial structure.

The urban spatial structure refers to the relationships arising from the underlying interaction of people, freight and information in an urban space [9]. The urban subcenter is the location wherein human social and economic activities are clustered, and since it impacts urban spatial reconstruction [10], it can represent a structural element of an urban sub-system within a metropolitan configuration [11].

According to the published literature [6], [12]–[15], a reasonable urban subcenter is defined as a site characterized by a higher employment density relative to adjacent locations while maintaining a significant effect on the overall employment density function. This indicates that the urban subcenter should have a significant effect both locally as well as globally.

Urban researchers have been concentrating on developing robust methods in order to identify urban subcenters, especially because the detailed definitions of their characteristics may vary.

Numerous studies regarding the identification of urban subcenters have produced fruitful results after years of research. McDonald [16] proposed an empirical method for identifying subcenters wherein a census zone is considered to be a subcenter if its employment density is larger than that of a neighboring zone. On the basis of McDonald's identification method, Giuliano and Small [17] suggested that a subcenter should constitute a contiguous set of zones, wherein the density of each zone is at least 10 per acre and the total number of employees is at least 10,000. Small and Song [18] utilized the same method with higher cutoffs, for which the least number

per acre was raised to 20, and the minimum number of employees was 20,000, and reasonable results were obtained. All of the above techniques are known as minimum cutoff point of density identification methods [3]. The primary shortcoming of this method is that the cutoff points must be obtained through trial and error, which requires researchers to thoroughly obtain sufficient local knowledge [6]. To reduce the dependency on this local knowledge, a two-stage nonparametric approach was initially proposed by. McMillen and McDonald [19] and subsequently modified by McMillen [6], McMillen and Lester [14], and McMillen and Smith [5]. The first step of this two-stage nonparametric approach employs locally weighted regression (a nonparametric estimator) to smooth the employment density as well as the sites which have significant positive residuals (local effects) that are considered to be subcenter candidates. The second step utilizes a semiparametric regression procedure to identify whether the subcenter candidates have a significant effect on the employment density (global effects).

Despite the broad utilization and advancement of minimum cutoff-point methods and two-stage nonparametric approaches in the study of urban spatial structure, some limitations still need to be eliminated:

1. The cutoff-points are unstable and require an excess of local knowledge; for example, McMillen and McDonald [20] changed the cutoff points by raising the minimum number per acre to 20 and the total number of employees to 20,000 because the smaller cutoff points applied during an identification of the Chicago subcenter led to an oversized and improper subcenter.

2. Employment data cannot be mobile in the urban subcenter. In fact, urban subcenters should represent sites that have strong internal associations but relatively weaker connections between internal and external regions. Consequently, it is not sufficient to identify urban subcenters solely using employment density.

3. Both of the minimum cutoff-point methods and the two-stage nonparametric approaches result in the inefficient identification of the spatial relationships between census tracts.

4. Since we cannot define a CBD precisely, it is difficult to confirm the most suitable density function in order to evaluate census data.

5. Algorithms do not account for the influence area of the subcenter; rather, they simply differentiate sites as either significant or insignificant. For a census tract, we are unable to constrain the subcenter that influences the tract most significantly.

Motivated by these limitations, this article introduces an alternative method to identify urban subcenters. A radiation model, a clustering algorithm and a spatial scan statistic are applied in this method. The radiation model is used to estimate commuting fluxes between each census tract, while the clustering algorithm is used to classify census tracts into different clusters, and the spatial scan statistic is used to identify ''hot-spots'' in each cluster as urban subcenters.

We apply this method to census data from 2010 and 2015 for the city of Wuhan, the results of which confirm that our proposed method is effective and can be used to analyze urban structures in future research.

The paper is organized as follows hereinafter.

In section 2, we introduce the theoretical background of our method, as well as the manner in which it operates.

We subsequently identify urban subcenters by applying our new method to the Wuhan census data for 2010 and 2015 in section 3.

Finally, we summarize and discuss our results in section 4.

II. METHODOLOGY

A. APPROXIMATION OF COMMUTING FLUXES

A distinct way in which to measure the mobility from one location to another is to count the physical interactions between those two locations. Zhong *et al.* [21] employed smart card data, which records all movement within the public transport systems in order to identify the Hubs, Centers and Borders in Singapore. Meanwhile, Zheng *et al.* [22] used GPS logs to infer people's modes of transport. In general, datasets that can represent physical interactions including the following: GPS-tracked vehicle datasets, telephone datasets, and public transportation datasets, among others. However, in some situations, these datasets are either inaccessible or unreliable. Thus, census tracts represent the most useful datasets with which to analyze urban spatial structure, and it is thus important to evaluate commuting fluxes between two areas using the census dataset relative to other auxiliary datasets.

Simini *et al.* [23] proposed a universal radiation model with which to predict population movement. It assumes that the commuting fluxes between an origin location (m) to a destination location (n) are dependent only upon the population of $m(m_i)$ and $n(n_j)$ and the population (s_{ij}) in a circle of radius r_{ij} (the distance from m to n) centered at m (excluding m_i and n_j). The characteristics of the radiation model overcome the drawbacks of the gravity model [24], [25], that is, lacking a rigorous derivation, lacking theoretical guidance, requiring excess parameters to fit the empirical data, requiring previous ancillary data, and so on.

Masucci *et al.* [26] assessed the robustness, universality and accuracy of the radiation model in addition to the gravity model by using three datasets (macroscopic datasets, small-scale datasets and moderate-sized datasets). The results obtained from the assessment indicate that the principles of the radiation model are reliable at large distances and for small to moderate destination population scales, although additional research is needed in order to improve the reliability of the model.

The radiation model is defined as follows:

$$F_{ij} = F_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})} \tag{1}$$

where m_i is the population of the origin area (m), and n_j is the population of the destination area (n). The population in the circle, whose center is the origin (that is, we use the centroid

of the origin location as the center) and whose radius is the Euclidean distance from the centroid of the origin area to the centroid of the destination area, is s_{ij} (excluding m_i and n_j).

The variable F_i in the model is proportional to the population m and is defined in equation (2), where N_c represents the total number of travellers and N is the total number of the population in the entire study area.

$$F_i = m_i \left(\frac{N_c}{N} \right) \tag{2}$$

Following the definition of F_i, equation (1) can be transformed into equation (3) as follows:

$$F_{ij} = \left(\frac{N_c}{N} \right) \frac{m_i^2 n_j}{\left(m_i + s_{ij}\right)\left(m_i + n_j + s_{ij}\right)} \tag{3}$$

$$C = \frac{N_c}{N} \tag{4}$$

following which

$$F_{ij} = C \frac{m_i^2 n_j}{\left(m_i + s_{ij}\right)\left(m_i + n_j + s_{ij}\right)} \tag{5}$$

Because N_c and N are constants in the study area, C is constant in equation (5). Afterwards, we calculate the degree of travelling volume between any two areas without multiplication by C (where C is an invariant constant in the equation).

In this study, we calculate the degree of commuting fluxes through the degree of physical interactions between any two locations using equation (3) (without multiplication by C), the radiation model for which is shown in Figure 1.

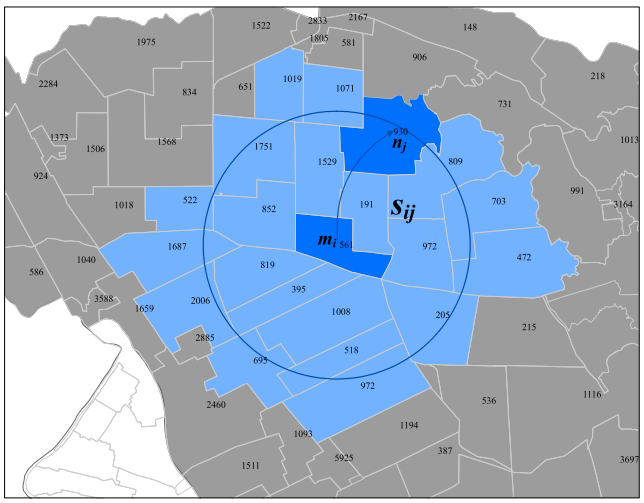

FIGURE 1. The radiation model employed in this study.

Note: The commuting fluxes can be properly estimated using the radiation model. The numbers shown in Figure 1 represent the population in each area, wherein the population of the origin tract (m_i) is 561 and the population of the destination tract (n_j) is 930. S_{ij} is composed of the population of tracts that are lightly colored but does not consist of m_i and n_j.

B. CLUSTERING BY COMMUTING FLUXES

A polycentric city should contain one or more subcenters in addition to the CBD. Boundaries need to be generated through the commuting fluxes in order to illustrate similar activities across parts of the city. These boundaries then partition the city into parts, after which we allocate them into different clusters.

Rodriguez and Laio [27] proposed an algorithm with which to classify points into different clusters, wherein the local density of the cluster center must be higher than the density of those in adjacent regions, and must have a relatively large distance from other cluster centers.

Two quantities (the local density ρ_i and distance δ_i) are computed for each point, the definitions for which are respectively shown in equations (6)-(7):

$$\rho_i = \sum_j \chi \left(d_{ij} - d_c \right) \tag{6}$$

$$\delta_i = \min_{j:\rho_j > \rho_i} \left(d_{ij} \right) \tag{7}$$

where χ (i) is the function of the point set, d_{ij} is the Euclidian distance between point i and point j, and d_c is a cutoff distance (Rodriguez and Laio suggested that d_c can be chosen such that the average number of neighbors is approximately 1% to 2% of the total number of points in the data set). If $d_{ij} > d_c$, then $\chi_x = 1$; otherwise, $\chi_x = 0$. This means that the local density ρ_i represents the number of points that are closer than the value of d_c to point i.

δ_i denotes the minimum distance between point i and any other point with a higher local density. For the point with the maximum local density point, δ_i is $max\left(d_{ij}\right)$.

After the local density and the relative distance have been computed for all of the points, the cluster centers are delineated as those points with a high local density and a large relative distance, that is, a high ρ and high δ. Those points with high δ but low ρ are considered to be outliers.

The classifying algorithm of Rodriguez and Laio [27] is performed in a single step, and can recognize distinct clusters regardless of their shape or the dimensionality of the space in which they are embedded. However, there are still some shortcomings in this algorithm. For example, while the results of the algorithm are robust with respect to the choice of d_c, it is difficult to derive a reasonable criterion with which to select the ''optimal'' d_c from both the theoretical and practical points of view [28].

To allocate census tracts into clusters, we define a dataset to represent the census tracts. For each point in the dataset, the location is equivalent with the centroid of the census tract, and the attribute corresponds to the census tract's population. Contrary to the work of Rodriguez and Laio [27], which only considers the position of a point regardless of the weight of the point, our study fixes the weight of each point as the population attribute of the point. This means that points with numerous surrounding points (that is, with low populations) may have lower local densities than those points with fewer surrounding points (with high populations).

To overcome these shortcoming regarding the arbitrary selection of d_c and to ensure the classifying algorithm has the ability to analyze census points (related not only to the density of surrounding points but to the population of the census area), we revise equation (6) using the radiation model [23].

After revision according to the radiation model, the local density $\rho_i^{'}$ is shown in equation (8):

$$\rho_i' = \sum_{j \neq i} F_{ji}' \tag{8}$$

where $F_{ji}^{'}$ is defined as equation (9):

$$F_{ji}' = \frac{m_j^2 n_i}{\left(m_j + s_{ji}\right)\left(m_j + n_i + s_{ji}\right)} \tag{9}$$

Therefore, the local density $\rho_i^{'}$ denotes the summation of the commuting fluxes with other census tracts. These commuting fluxes are calculated using the revised radiation model. Figure 2 shows an example of the local density of the 150$^{\text{th}}$ tract.

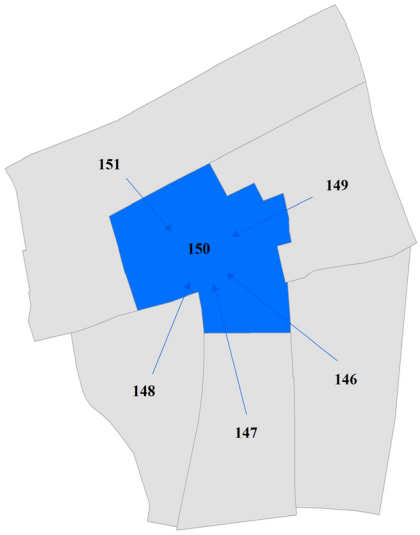

FIGURE 2. The local density of the census tract.

Note: The local density of the 150$^{\text{th}}$ tract is the summation of the commuting fluxes with other tracts (the 146$^{\text{th}}$, 147$^{\text{th}}$, 148$^{\text{th}}$, 149$^{\text{th}}$, and 151$^{\text{st}}$ tracts). Mathematically, $\rho_{150}^{'} = F_{146,150}^{'} + F_{147,150}^{'} + F_{148,150}^{'} + F_{149,150}^{'} + F_{151,150}^{'}$.

After obtaining the local density, we classify each census tract into a different cluster using the algorithm of Rodriguez and Laio [27].

The distance $\delta_i^{'}$ is similar to equation (7) and the centroid point that has the highest local density leads to $\delta_i^{'}$ being equal to $\max(d_{ij})$.

Subsequently, the density peak of each census tract can be identified by the local density $\rho_i^{'}$ and the distance $\delta_i^{'}$. From equation (7) and equation (8), the revised algorithm is a non-parameter algorithm that only depends on the population of the census tracts and the distance between the centroid points of the census tracts.

It is apparent that equations (8)-(9) are used to simulate the 'physical interactions' between each census tract. If datasets of 'physical interactions' are available and reliable, these datasets can be applied directly towards classifying census tracts into clusters without a 'physical interaction simulation'.

C. IDENTIFICATION OF SUBCENTERS USING A SPATIAL SCAN STATISTIC

Spatial scan statistics [29]–[31] and spatial autocorrelation indices [32], [33] (Local Moran's I and the Getis statistic, among others) are widely utilized to detect "hot spots" for agriculture, air pollution, drug offences, and so on [34]–[36].

There is a major shortcoming in the use of spatial autocorrelation indices such that although many spatial weights matrices have been created (e.g., the inverse distance raised to some power, k nearest neighbors, centroid points within a predefined threshold d_θ, ranked distances, etc.), some models that employ these schemes are misspecified.

Unlike spatial autocorrelation indices, spatial scan statistics are based on a maximum likelihood ratio test, and employ a scanning window to detect "hot-spots" in a given study area [31].

Most spatial scan statistics are based upon special distributions, and may lead to very different results. Thus, we adopt a distribution-free spatial scan statistic in order to discover the center of each cluster.

The distribution-free spatial scan statistic [37] has a simple criterion: the mean value inside the center should be significantly higher than the mean value outside of the center.

Let $D_1, D_2, \ldots, D_i$ denote the local densities associated with the centroids of the census tracts. Assume for each D_i that

$$E\left(D_i\right) = \mu \text{ and } V\left(D_i\right) = \sigma^2 \text{ and } Cov\left(D_i, D_j\right) = 0$$

Then, for the center Z as well as the area outside the center $\bar{Z}$, the difference of the mean values is as follows:

$$D(Z) = \mu(Z) - \mu(\overline{Z}) \tag{10}$$

$\mu\left(Z\right)$ is defined as follows:

$$\mu\left(Z\right) = \frac{\sum_{i=1}^{n} D_i \left(s_i \in Z\right)}{\sum_{i=1}^{n} 1 \left(s_i \in Z\right)} \tag{11}$$

where s_i is the centroid of the i^{th} census tract, and D_i is the local density of s_i.

The variance of $D\left(Z\right)$ is:

$$V(D(Z)) = \sigma^2\left(\frac{1}{n(Z)} + \frac{1}{n(\overline{Z})}\right) \tag{12}$$

where $n\left(Z\right)$ denotes the number of census tracts inside the center Z; thus,

$$I(Z) = \frac{\sqrt{n(Z)n(\overline{Z})}}{\sqrt{n(Z) + n(\overline{Z})}}\left[\mu(Z) - \mu(\overline{Z})\right] \tag{13}$$

is constructed as a concentration index λ_p in order to detect the tract center wherein

$$\lambda_p = \max(I(Z)) \tag{14}$$

To employ the concentration index, a scanning window is needed. Most spatial scan statistics use scanning windows of predefined shapes (circular or elliptic) in order to detect "hot-spots". In this paper, a circular scanning window is used, the size of which is variable, and the number of areas does not exceed 50% of the number of areas within the study region. Figure 3 shows the circular scanning windows with different sizes and the scanning results.

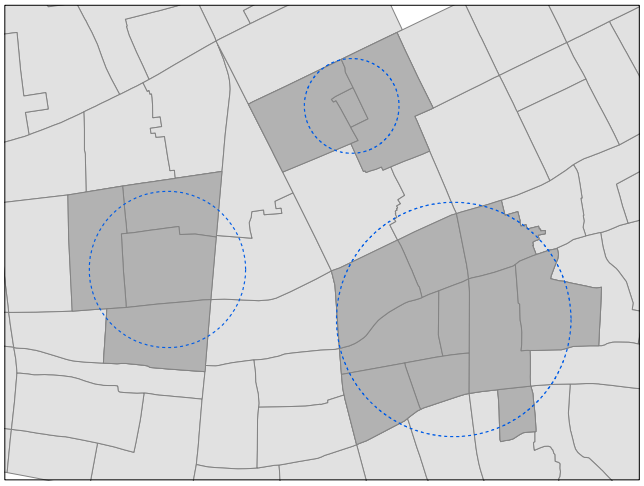

FIGURE 3. Census tracts within scanning windows.

Note: The centroid of the census tracts are within the scanning windows. If part of a census tract is within the scanning window but the centroid is not, then it should be excluded from the scanning results.

We calculate the concentration index for each scanning result, and the region whose concentration index value is the highest within a cluster is the most likely subcenter.

After the most likely subcenter has been identified, a *p-value* is used to test whether it is significant. Here, we define the *p-value* in equation (15):

$$p_\lambda = \frac{\sum_{n=1}^{N} 1(\lambda^n > \lambda)}{N+1} \tag{15}$$

where N is the number of simulated datasets and λ^n is the result (the max concentration index) of the n^{th} simulation. For each simulation, the population is randomly reallocated among the census tracts.

In this paper, N is set as 999 and the *confidence level* is set as 95%, which means the *p-value* of a subcenter should be less than 0.05 if it is significant.

The above subsections demonstrate that our method overcomes limitations one through five. These limitations existed in both the minimum cutoff point of density identification method as well as the two-stage nonparametric approach.

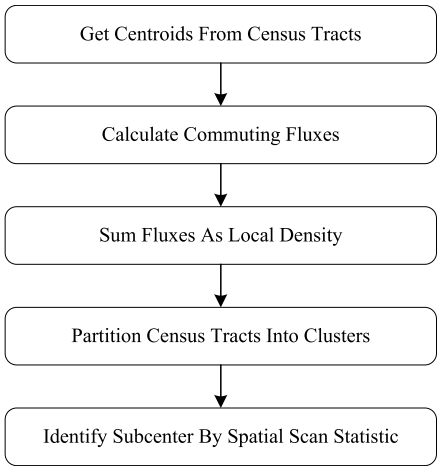

FIGURE 4. Work-flow diagram of our proposed method.

Our approach does not require prior local knowledge because it uses the radiation model to calculate commuting fluxes, which are then used to reveal the physical interactions between census tracts and to classify the census tracts into clusters. These clusters are then utilized to represent the area of influence of each urban subcenter.

In addition, our approach does not incorporate a density estimating function. Figure 4 illustrates the work-flow of our proposed method.

III. APPLICATION TO WUHAN, CHINA

A. DATASETS

The Wuhan urban developing area, which is the area examined in this study, is located in Hubei Province, China. As the largest and most important city of Central China, Wuhan has a total administrative area of approximately 8,494 km^2 and a population exceeding 10,338,000. The Wuhan Metropolitan region includes the urban area and its adjacent outer suburbs. The urban area contains 7 districts (Jiangan, Jianghan, Qiaokou, Hanyang, Wuchang, Qingshan and Hongshan), and the suburban area contains 6 districts (Caidian, Jiangxia, Dongxihu, Hannan, Huangpi and Xinzhou). The distribution of the population of Wuhan is imbalanced and the spatial differentiation of population growth is large. While the population growth in the suburbs is not significant, population growth is more dramatic in urban areas.

Wuhan's urban developing area is comprised of the 13 aforementioned districts: Jiangan, Jianghan, Qiaokou, Hanyang, Wuchang, Qingshan, Hongshan, Caidian, Jiangxia, Dongxihu, Hannan, Huangpi and Xinzhou. The first 7 districts are defined as urban areas and constitute a greater number of census tracts, which are smaller in size than those of the latter six districts. Here, we have obtained the centroid for each of the census tracts, which are plotted in Figure 5. From the scatter diagram, we can observe that the census tracts in the urban areas are much smaller than the tracts in the suburban areas. Therefore, urban census tracts can provide more precision than suburban census tracts. Table 1 provides

TABLE 1. Summary of statistics of the census data.

	Maximum	Minimum	Mean	Standard Deviation	Total
Area (km^2) 2010	18.9	0.013	1.49	2.2	2,721
Population (persons)	55,863	6	4,436	4,976	8,056,770
Density (persons/km^2)	284,507	3.43	21,118	28,419	—
2015					
Population (persons)	66,125	14	4,599	5,323	8,353,457
Density (persons/km^2)	307,782	8.11	22,290	30,830	—

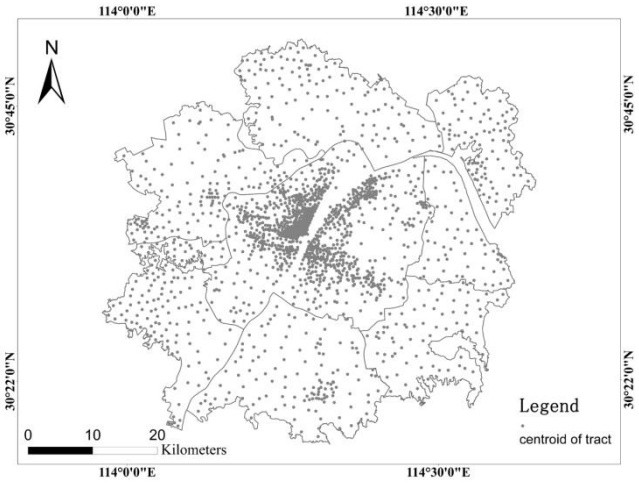

FIGURE 5. Locations of the census tracts for Wuhan.

a summary of the statistics for the census data used in this paper.

B. CLUSTERS OF THE CENSUS TRACTS IN THE STUDY REGION

As mentioned above, in order to make use of the clustering algorithm, we conduct an analysis using the centroids of census tracts instead of the original census data. We exclude some centroids due to zero population in order to avoid division by zero. In a strict sense, the local density of a centroid is needed to calculate commuting fluxes with other centroids, and to calculate the total. However, this procedure is rather time-consuming with a time complexity of $O(n^2)$. We conclude through experiments that, if the distance between two centroids is greater than 5 km, the value of the commuting flux between either centroid is so small that it can be omitted. Thus, after we obtain the local densities of the centroids, only centroids within 5 km are considered.

Figure 6(a) illustrates that the plot of $\rho_i\delta_i$ displays in descending order for the 2010 census tracts of Wuhan, while Figure 6(b) shows the plot of δ_i as a function of ρ_i. The highest value of $\rho_i\delta_i$ in Figure 6(a) is excluded in order to enhance the readability of the graph, after which there are 5 other

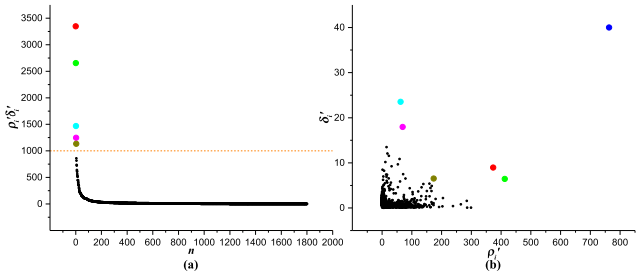

FIGURE 6. Results for the 2010 census tracts of Wuhan. (a) The values of $\rho_i\delta_i$ from (b) in descending order. The highest value of $\rho_i\delta_i$, the blue point in (b), is excluded in order to increase the readability of the graph. (b) Density-distance plot for the census tracts.

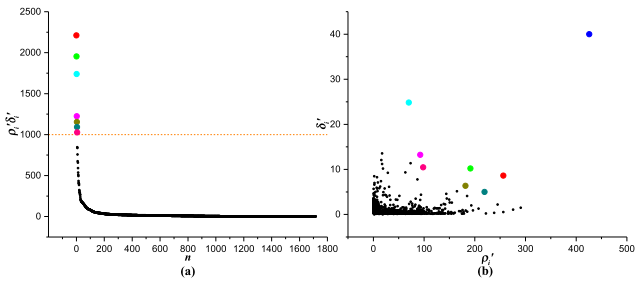

FIGURE 7. Results for the 2015 census tracts of Wuhan.

points with relatively high values that correspond to points in Figure 6(b) by color. A total of 6 points with different colors have both relatively high values of the local density ρ_i and of the distance δ_i. Following the population growth and a change of the spatial distribution, the plots of δ_i and of $\rho_i\delta_i$ for the 2015 Wuhan census tracts in Figure 7 exhibit different results from those of the 2010 census data. The number of points with relatively high values of both the local density ρ_i and of the distance δ_i increase to 8, but demonstrate an overall smaller value gap for $\rho_i\delta_i$.

A comparison between the 2010 and 2015 census tracts for Wuhan reveals an overall change of the spatial distribution. The number of clusters of Wuhan census tracts changed from 6 for 2010 to 8 for 2015. Although the quantity increased, their density values decreased, which indicates that the spatial distribution has dispersed.

TABLE 2. Comparison of analyses using census tracts from 2010 and 2015 for Wuhan.

Year of census tracts	2010	2015
Total Population	8,056,595	8,352,633
Average Population	4,436	4,599
Number of Clusters	6	8
Average Flux	26,045	26,695
Minimum Flux	0	0
Maximum Flux	763,008	425,912

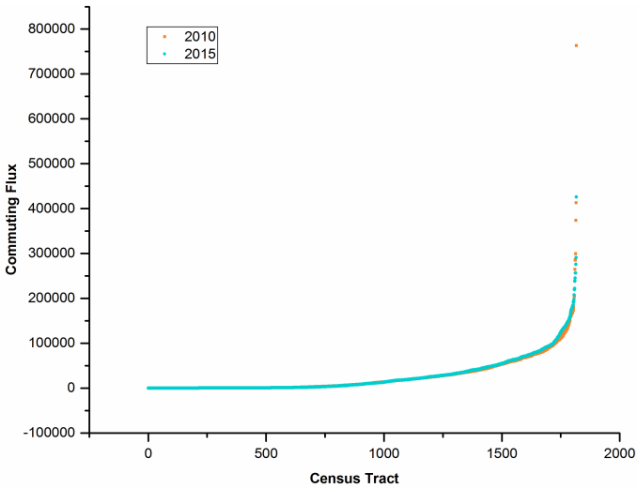

FIGURE 8. The commuting flux of the census tracts in 2010 and 2015.

Table 2 illustrates the overall properties of the census tracts for both 2010 and 2015. We can recognize that the total population and average commuting fluxes of each census tract in the study area increased gradually, and the number of clusters increased by 2 from 2010 to 2015. In addition, the maximum commuting flux in 2015 was much lower than in 2010.

Another scatter diagram is shown in Figure 8, which illustrates the commuting flux in ascending order for the census tracts from 2010 and 2015. It is apparent therein that, although the maximum commuting flux in 2010 is much higher than in 2015, the average commuting flux in 2010 is generally lower than in 2015. This verifies our conclusion that the spatial distribution throughout Wuhan has dispersed.

From Figure 6 to Figure 8 and from Table 2, we find that the city of Wuhan is becoming more polycentric. Although the traditional subcenters still play important roles, new subcenters obviously weaken their influence. This phenomenon can be quantified by the decrease of $\rho_i\delta_i$ values: the maximum value of $\rho_i\delta_i$ in 2010 and 2015 is 30,520,320 and 17,036,480, respectively.

After the cluster centers have been identified, each census tract is allocated to the nearest census tract centroid with the higher local density. To make this procedure complete more quickly, these census tracts are sorted by local density in descending order, the results of which are shown in Figure 9 and Figure 10 for 2010 and 2015, respectively.

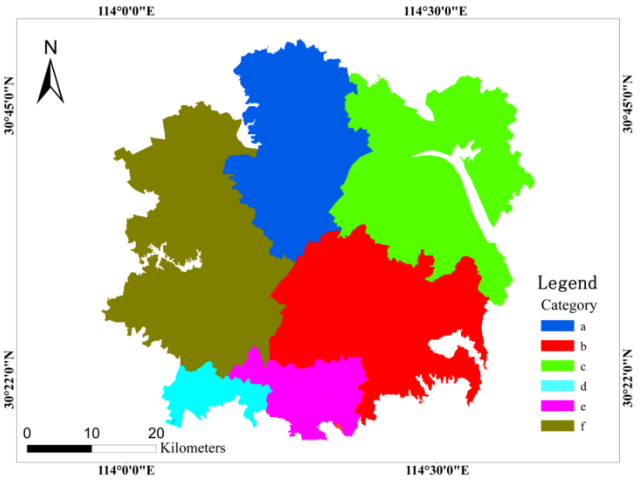

FIGURE 9. Clusters detected from the 2010 Wuhan census data.

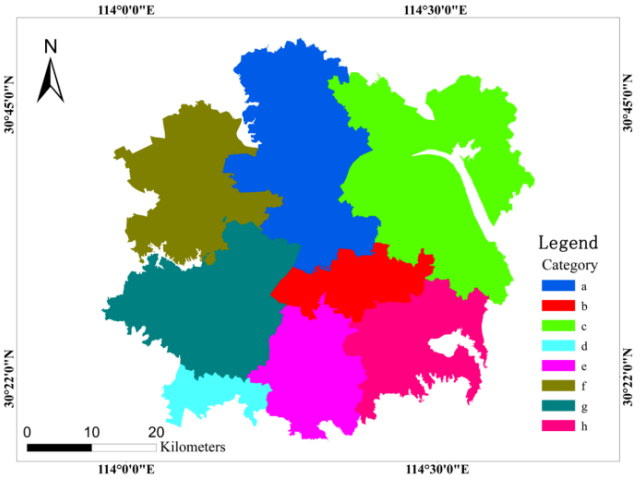

FIGURE 10. Clusters detected from the 2015 Wuhan census data.

TABLE 3. The maximum commuting flux values among the different clusters.

Cluster	2010	2015
A	763,008	425,912
B	373,533	256,415
C	412,651	191,595
D	62,349	70,048
E	69,378	92,697
F	173,606	181,675
G	—	219,451
H	—	98,330

Although we classify the clusters according to decreasing $\rho_i\delta_i$ values, it is not implied that the commuting flux is also descending. In fact, some clusters are characterized by low-commuting flux values but have a high relative distance value. Table 3 shows the maximum commuting flux values of the clusters obtained using the 2010 and 2015 census data.

The degree of difference in the $\rho_i\delta_i$ values and commuting fluxes indicates that there are two types of subcenters.

TABLE 4. Statistics for the subcenters in 2010.

Subcenter Cluster	Average Flux	λ_p	p-value
A	66,699	564,325	0.001
B	349,431	373,533	0.019
C	43,949	303,722	0.001
D	48,648	63,696	0.002
E	27,982	76,680	0.01
F	449,463	754,571	0.003

Subcenters with high commuting fluxes have some subcenters nearby, but it has more of an influence on nearby tracts. Subcenters with low commuting fluxes are not adjacent to any subcenters for a long distance. Although it does not have as great of an influence on neighboring tracts as subcenters with high commuting fluxes, subcenters with low commuting fluxes are the most important sites over a broad region.

Comparing the variations in the clusters of census tract between 2010 and 2015, the greatest changes occurred in the west and south of the study area. A 2010 cluster in the west (f) was divided into two clusters (f and g) in 2015. Two 2010 clusters in the south (b and e) were merged and then re-divided into three clusters in 2015 (b, e and h). The other clusters appear to have remained relatively stable.

It is observed that the new cluster in the west (g) corresponds to the Hankou railway station and its adjacent areas and that the new cluster in the south (h) corresponds to Guanggu square and its neighboring areas. Both the Hankou railway station and Guanggu square have rapidly developed over these years.

C. SUBCENTERS IN THE STUDY REGION

For each cluster, census tracts are selected using a distribution-free (DB-free) spatial scan statistic because the distribution of census tracts is untractable. In this study, the *p-value* is set as 5%. We then generate 999 simulated datasets to represent the power of the DB-free spatial scan statistic. For each simulation, we calculate the concentration index (λ_p) and determine whether it is greater than the value of λ_p calculated using the census tracts in a cluster. The results are shown in Figure 11 and Figure 12 for 2010 and 2015, respectively.

From Figure 11, we can see that there are 6 subcenters, while 8 are observed from Figure 12. This suggests that all of the subcenters detected using the DB-free spatial scan statistic have a p-value lower than 5%. We also find from Figure 11 and Figure 12 that the distribution of the subcenters transformed from concentrated to dispersed. In other words, Wuhan became more polycentric from 2010 to 2015. Statistics for the subcenters in 2010 and 2015 can be seen in Table 4 and Table 5, respectively.

After the subcenters have been identified, we try to determine whether the subcenters have a relationship with the locations of metro stations. Because traffic data is not publicly available, we obtain data from the official microblog (http://weibo.com/u/3186945861) of METRO, WUHAN.

Every day, the official microblog posts data regarding the Top 5 metro stations with respect to traffic flow. As a consequence, we can use these data to locate 'hot' metro stations. Ultimately, 428 data records were obtained from the microblog.

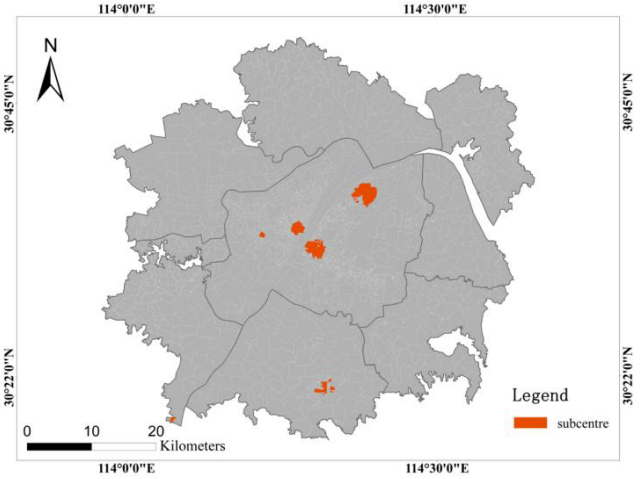

FIGURE 11. The location of subcenters in Wuhan, 2010.

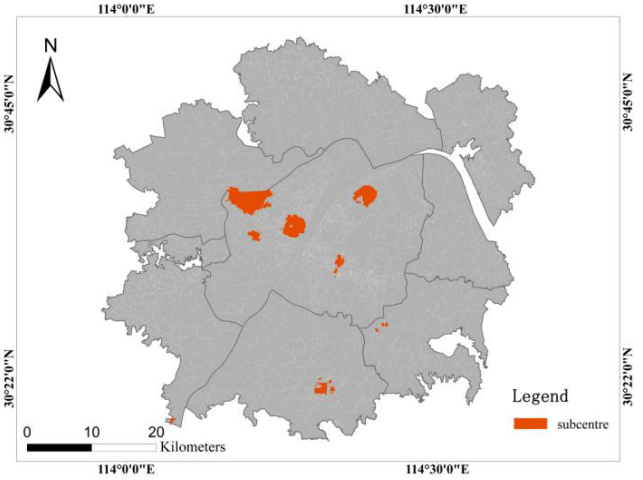

FIGURE 12. The location of subcenters in Wuhan, 2015.

TABLE 5. Statistics for the subcenters in 2015

Subcenter Cluster	Average Flux	λ_p	p-value
A	89,555	587,243	0.008
B	256,415	228,936	0.033
C	45,209	309,908	0.008
D	57,061	74,566	0.003
E	92,697	84,416	0.001
F	47,345	228,479	0.016
G	48,127	276,271	0.022
H	98,330	93,286	0.005

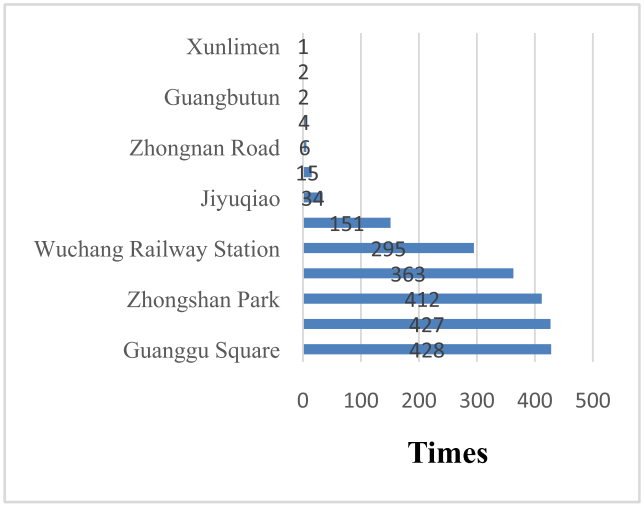

FIGURE 13. The 'hot' metro stations.

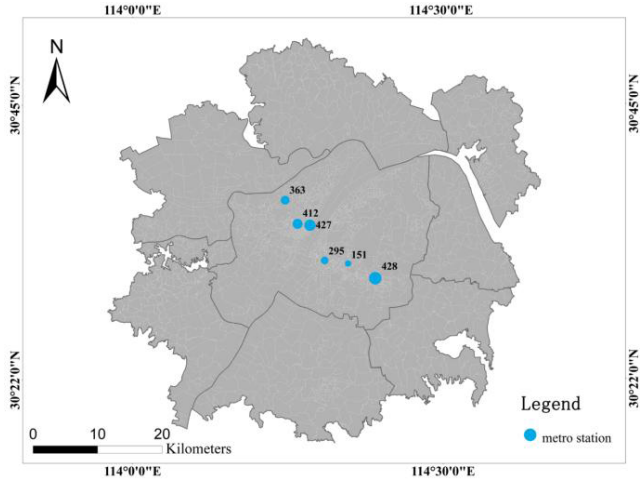

FIGURE 14. The 6 metro stations with the largest traffic flow values.

Figure 13 shows the statistical results for 'hot' metro stations, while Figure 14 illustrates the locations of the 6 metro stations with the highest measured traffic flow (determined using over 100 days' worth of data).

We discover that 4 of the metro stations have the same location with subcenters in 2015, except for Guanggu Square and Wuchang Railway Station. Subcenters in cluster A contain two metro stations: Jianghan Road and Zhongshan Park. With the exception of the top 6 metro stations, we compare the locations of subcenters with other metro stations not included in Figure 14. We subsequently find that all of the subcenters in urban areas contain at least one metro station. This is the same conclusion reached by Zhong *et al.* [21].

D. COMPARISONS OF METHODOLOGIES

1) COMPARISON WITH THE CUTOFF-POINT APPROACH AND THE TWO-STAGE APPROACH

The minimum cutoff point of density identification method and the two-stage nonparametric approach are based on the same exact idea wherein a subcenter should exert a significant influence on both the local and global scope. The first step in the two-stage nonparametric approach is to use Z-scores to identify candidate subcenter sites. In most literature, the Z-score is defined as 1.28 or 1.96, respectively corresponding to a 5% or 10% significance level. The second stage is based on the candidate subcenter sites, which are the sites that have statistically significant effects on a global scale.

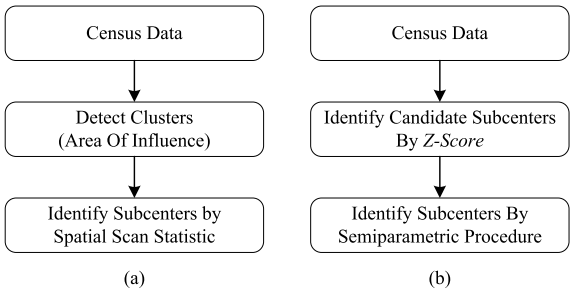

FIGURE 15. Differences between the two approaches. (a) Our approach, and (b) the two-stage nonparametric approach.

Our research could be regarded as a reversed two-stage approach. Unlike the two-stage nonparametric approach, our approach conducts an analysis at a global scale, following which the real subcenters are obtained.

The differences between two approaches are shown in Figure 15.

2) COMPARISON WITH MULTIPLE CLUSTER SPATIAL SCAN STATISTIC

More than one significant cluster may exist within a geographical region, but only if their *p-values* are less than 5%. Many methods have been proposed to retrieve multiple clusters in a study region. Zhang *et al.* [38] proposed a sequential version of the spatial scan statistic to detect other significant clusters, except for the most likely cluster (MLC), by removing the areas which comprise the MLC as well as the areas adjacent to the MLC, or by making the ratio between the number of cases and the population within the MLC equivalent to the ratio outside of it.

Contrary to the method of Zhang *et al.* [38], we divided the census tracts into smaller clusters so the chance that multiple significant "hot-spots" coexist could be reduced for each cluster. Therefore, we only obtained the MLC for each group.

IV. CONCLUSIONS AND DISCUSSIONS

This paper introduces an alternative method with which to identify urban subcenters. These subcenters are then used to analyze the variation of urban spatial structure in 2010 and 2015 in the city of Wuhan, China. Unlike previous research, our approach does not require arbitrary cutoff values and prior local knowledge, and it is entirely parameter-free. Some methods have used spatial network analyses to identify the spatial structure of city hubs, centers and borders utilizing 'big' datasets sourced from the automatic smart card fare

collection system. In many other cases, however, no such sets of data or ancillary data exist, and the only complete and effective datasets are census datasets. Therefore, such methodologies cannot be widely applied. We have demonstrated that our approach is advantageous for the identification of urban subcenters and the detection of urban spatial structure variation.

The method introduced in this paper is discussed from three different perspectives: the calculation of commuting fluxes, the classification of clusters and the identification of subcenters. The calculation of commuting fluxes only requires census data. For each census tract, we calculate the commuting fluxes between it and other census tracts. The summation of commuting fluxes is then used to represent the local density of the census tracts. This avoids the choice of arbitrary cutoff values of d_c. In fact, our method can be applied to data sets insomuch that the interactions between each tract can be calculated. Subcenters and regions of influence are key factors with which to analyze and demonstrate variations in urban spatial distribution and urban spatial structure movement. Our approach has focused on these vital factors and accordingly applied them to the 2010 and 2015 Wuhan census data. The results demonstrate that there were some significant changes in west and south of the study area and that the city became relatively polycentric by the year 2015.

There is still a great deal of work that must be done in order to improve our approach. In our method, a radiation model was used to calculate the commuting fluxes between each census tract with the others. We then calculated the local density by summing up commuting fluxes. Although the reliability of the radiation model at large distances, as well as for small and moderate destination population scales, has been assessed, additional research is required to improve the reliability of the model. For example, in some extreme cases, the radiation model may generate large errors and lead to some results that are incorrect. The first thing needed in order to improve our research is to increase the accuracy of the radiation model to ensure improved local density results. We filtered local density peak points by sorting the values of ρ_i and δ_i, following which any value larger than 1000 was chosen as local density peak point. This cutoff is arbitrary and may lead to different results if we select a different value. Although the published literature has introduced an automatic selection technique for local density peak points, it is only suitable for local density peak points that have strong, large values of local density in addition to relative distances, which can omit other weaker points. Developing a method with which to choose these local density peak points accurately is another key factor for improving our approach. A spatial scan statistic was used to detect subcenters in each cluster. Because the distribution of data sets is untraceable, we adopt a distribution-free spatial scan statistic in this paper. However, this distribution-free method can be unstable if the real subcenter is not circular. Although an irregularly shaped scan window can be used to detect subcenters, a few inherent issues need to be resolved. Although the pre-grouping procedure can reduce the chance of coexisting multiple significant ''hot-spots'', it cannot eliminate the potential. Moreover, if there is more than one significant ''hot-spot'' in one cluster, should the MLC be left as the sole subcenter in that cluster, or should the cluster be divided into smaller clusters, which can then be chosen as subcenters?

Future work should consequently focus on the following aspects: the accuracy of the calculations for commuting fluxes, the rationality of the selection of local density peak points, and the stability of the spatial scan statistic. These improvements will contribute toward a better understanding of the urban spatial distribution and the urban spatial structure.

REFERENCES

[1] M.-Á. Garcia-López, ''Population suburbanization in Barcelona, 1991Ű2005: Is its spatial structure changing?'' *J. Housing Econ.*, vol. 19, no. 2, pp. 119–132, 2010.
[2] A. Anas, R. Arnott, and K. A. Small, ''Urban spatial structure,'' *J. Econ. Literature*, vol. 36, no. 3, pp. 1426–1464, 1998.
[3] Q. Pan and L. Ma, ''Employment subcenter identification: A GIS-based method,'' *Texas Southern Univ. Sci. Urban Econ.*, vol. 21, no. 2, pp. 63–82, 2006.
[4] R. Louf and M. Barthelemy, ''Modeling the polycentric transition of cities,'' *Phys. Rev. Lett.*, vol. 111, no. 19, p. 198702, 2013.
[5] D. P. McMillen and S. C. Smith, ''The number of subcenters in large urban areas,'' *J. Urban Econ.*, vol. 53, no. 3, pp. 321–338, 2003.
[6] D. P. McMillen, ''Nonparametric employment subcenter identification,'' *J. Urban Econ.*, vol. 50, no. 3, pp. 448–473, 2001.
[7] D. E. Bloom, D. Canning, and G. Fink ''Urbanization and the wealth of nations,'' *Science*, vol. 319, no. 3008, pp. 772–775, 2008.
[8] G. Grekousis and G. Mountrakis, ''Sustainable development under population pressure: Lessons from developed land consumption in the conterminous U.S,'' *Plos One*, vol. 10, no. 3, p. e0119675, 2015.
[9] J.-P. Rodrigue, *The Geography of Transport Systems*. Evanston, IL, USA: Routledge, 2013.
[10] P. Sun, C. Xiu, R. Pang, and W. Song, ''Improvement and application of the sectoral enterprises geographic clustering model and its formed urban structure,'' *J. Urban Planning Develop.*, vol. 142, no. 4, p. 05016006, 2016.
[11] J. Roca Cladera, C. R. M. Duarte, and M. Moix, ''Urban structure and polycentrism: Towards a redefinition of the sub-centre concept,'' *Urban Stud.*, vol. 46, no. 13, pp. 2841–2868, 2009.
[12] D. P. McMillen, ''Identifying sub-centres using contiguity matrices,'' *Urban Stud.*, vol. 40, no. 1, pp. 57–69, 2003.
[13] Y. Zhang and K. Sasaki, ''Effects of subcenter formation on urban spatial structure,'' *Regional Sci. Urban Econ.*, vol. 27, no. 3, pp. 297–324, 1997.
[14] D. P. McMillen and T. W. Lester, ''Evolving subcenters: Employment and population densities in Chicago, 1970-2020,'' *J. Housing Econ.*, vol. 12, no. 1, pp. 60–81, 2003.
[15] S. G. Craig and P. T. Ng, ''Using quantile smoothing splines to identify employment subcenters in a multicentric urban area,'' *J. Urban Econ.*, vol. 49, no. 1, pp. 100–120, 2001.
[16] J. F. McDonald, ''The identification of urban employment subcenters,'' *J. Urban Econ.*, vol. 21, no. 2, pp. 242–258, 1987.
[17] G. Giuliano and K. A. Small, ''Subcenters in the Los Angeles region,'' *Regional Sci. Urban Econ.*, vol. 21, no. 2, pp. 163–182, 1991.
[18] K. A. Small and S. Song, ''Population and employment densities: Structure and change,'' *J. Urban Econ.*, vol. 36, no. 3, pp. 292–313, 1994.
[19] D. P. McMillen and J. F. McDonald, ''A nonparametric analysis of employment density in a polycentric city,'' *J. Regional Sci.*, vol. 37, no. 4, pp. 591–612, 1997.
[20] D. P. McMillen and J. F. McDonald, ''Suburban subcenters and employment density in metropolitan Chicago,'' *J. Urban Econ.*, vol. 43, no. 2, pp. 157–180, 1998.
[21] C. Zhong, S. M. Arisona, X. Huang, M. Batty, and G. Schmit, ''Detecting the dynamics of urban structure through spatial network analysis,'' *Int. J. Geograph. Inf. Sci.*, vol. 28, no. 11, pp. 2178–2199, 2014.

[22] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, "Understanding mobility based on GPS data," in *Proc. Ubiquitous Comput. Int. Conf. (UBICOMP)*, Seoul, South Korea, Sep. 2008, pp. 312–321.
[23] F. Simini, M. C. González, A. Maritan, and A. L. Barabásim, "A universal model for mobility and migration patterns," *Nature*, vol. 484, no. 7392, pp. 96–100, 2012.
[24] S. Erlander and N. F. Stewart, *The Gravity Model in Transportation Analysis: Theory and Extensions*. Utrecht, The Netherlands: VSP, 1990.
[25] L. Mátyás, "Proper econometric specification of the gravity model," *World Econ.*, vol. 20, no. 3, pp. 363–368, 1997.
[26] A. P. Masucci, J. Serras, A. Johansson, and M. Batty, "Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 88, no. 2, p. 022812, 2013.
[27] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.
[28] X.-F. Wang and Y. Xu, "Fast clustering using adaptive density peak detection," *Statist. Methods Med. Res.*, vol. 2015, p. 0962280215609948, Oct. 2015. [Online]. Available: http://journals.sagepub.com/doi/abs/10.1177/0962280215609948
[29] I. Jung, M. Kulldorff, and O. J. Richard, "A spatial scan statistic for multinomial data," *Statist. Med.*, vol. 29, no. 18, pp. 1910–1918, 2010.
[30] L. Duczmal and R. Assunção, "A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters," *Comput. Statist. Data Anal.*, vol. 45, no. 2, pp. 269–286, 2004.
[31] M. Kulldorff, "A spatial scan statistic," *Commun. Statist. Theory Methods*, vol. 26, no. 6, pp. 1481–1496, 1997.
[32] C. Dormann *et al.*, "Methods to account for spatial autocorrelation in the analysis of species distributional data: A review," *Ecography*, vol. 30, no. 5, pp. 609–628, 2007.
[33] C. Fan and S. Myint, "A comparison of spatial autocorrelation indices and landscape metrics in measuring urban landscape fragmentation," *Landscape Urban Planning*, vol. 121, no. 1, pp. 117–128, 2014.
[34] M. Quick and J. Law, "Exploring hotspots of drug offences in Toronto: A comparison of four local spatial cluster detection methods," *Can. J. Criminol. Criminal Justice*, vol. 55, no. 2, pp. 215–238, 2013.
[35] I. J. Marasteanu, C. L. K. Liang, and S. Goetz, "Spatial and cluster analysis for multifunctional agriculture in New England Region," in *Proc. Annu. Meeting*, Minneapolis, Minnesota, Jul. 2014. [Online]. Available: http://econpapers.repec.org/paper/agsaaea14/170314.htm
[36] B. Zou, F. Peng, N. Wan, K. Mamady, and G. J. Wilson, "Spatial cluster detection of air pollution exposure inequities across the United States," *Plos One*, vol. 9, no. 3, p. e91917, 2014.
[37] L. Cucala, "A distribution-free spatial scan statistic for marked point processes," *Spatial Statist.*, vol. 10, pp. 117–125, Nov. 2014.
[38] Z. Zhang, R. Assunção, and M. Kulldorff, "Spatial scan statistics adjusted for multiple clusters," *J. Probab. Statist.*, vol. 2010, Jun. 2010, Art. no. 642379, doi:10.1155/2010/642379. [Online]. Available: https://www.hindawi.com/journals/jps/2010/642379/cta/

ZHONGLIANG FU received the B.S., M.S., and Ph.D. degrees in photogrammetry and remote sensing from the Wuhan Technical University of Survey and Mapping, Wuhan, China, in 1985, 1988, and 1996, respectively. He is currently a Professor and a Ph.D. Advisor with the School of Remote Sensing and Information Engineering, Wuhan University. He is also the Director of the Geographic Information System Department.

His interests include spatial data management and update, remote sensing image processing and analysis, map scanning image recognition, vehicle license plate recognition, and geographic information engineering technology.

KAICHUN ZHOU received the B.S. and M.S. degrees in geographic information system from Central South University, Changsha, China, in 2010 and 2013, respectively. He is currently pursuing the Ph.D. degree in cartography and geographical information engineering with the School of Remote Sensing and Information Engineering, Wuhan University. His research interests are spatial clustering, detection, and analysis.

LIANG FAN received the B.S. and M.S. degrees in remote sensing science and technology from Wuhan University, Wuhan, China, in 2013 and 2015, respectively, where he is currently pursuing the Ph.D. degree in photogrammetry and remote sensing with the School of Remote Sensing and Information Engineering. His research interests are spatial data management and update, multisource spatial data matching and fusion, and spatial data analysis.