

Article

Fast Identification of Urban Sprawl Based on K-Means Clustering with Population Density and Local Spatial Entropy

Lingbo Liu ¹ , Zhenghong Peng ², Hao Wu ^{2,*} , Hongzan Jiao ², Yang Yu ¹  and Jie Zhao ¹

¹ Department of Urban Planning, School of Urban Design, Wuhan University, Wuhan 430072, China; lingbo.liu@whu.edu.cn (L.L.); yuyang1@whu.edu.cn (Y.Y.); whuzhj@whu.edu.cn (J.Z.)

² Department of Graphics and Digital Technology, School of Urban Design, Wuhan University, Wuhan 430072, China; pengzhenghong@whu.edu.cn (Z.P.); Jiaohongzan@whu.edu.cn (H.J.)

* Correspondence: wh79@whu.edu.cn; Tel.: +86-27-6877-3062

Received: 28 June 2018; Accepted: 27 July 2018; Published: 31 July 2018



Abstract: As urban sprawl is proven to jeopardize the sustainability system of cities, the identification of urban sprawl is essential for urban studies. Compared with previous related studies which tend to utilize more and more complicated variables to recognize urban sprawl while still retaining an element of uncertainty, this paper instead proposes a simplified model to identify urban sprawl patterns. This is a working theory which is based on a diagram interpretation of the classic urban spatial structure patterns of the Chicago School. The method used in our study is K-means clustering with gridded population density and local spatial entropy. The results and comparison with open population data and mobile phone data verify the assumption and furthermore indicate that the accuracy of source population data will limit the precision of output identification. This article concludes that urban sprawl is mainly dominated by population and surrounding unevenness. Moreover, the Floating Catchment Area (FCA) local spatial entropy method presented in this research brings about an integration of Shannon entropy, Tobler's first law of geography and the Moore neighborhood, improving the spatial homogeneity and locality of Batty's Spatial Entropy model which can only be used in a general scope.

Keywords: urban sprawl; K-means clustering; Floating Catchment Area (FCA); local spatial entropy; population density; Elbow method

1. Introduction

Urban spatial expansion is described as a process in which the city encroaches surrounding open space and extends the territory of urban areas, due to its growing population, increasing incomes and decreasing commuting costs [1]. As a disfavored type of expansion, urban sprawl refers to excessive spatial growth of cities, which results in negative consequences such as rapid loss of farmland, despoiled ecological system and swelling traffic congestion in both the West and the East [2–10], which resulted in a wide range of economic [2], social [3], health [4] and environmental problems [11], jeopardizing the sustainability system of cities and becoming a major concern for urban studies. Urban sprawls have attracted considerable attention from various perspectives, such as its definition, measurement, cause, effect, simulation and solutions [12–18]; however, it still has no universal definition [19–21], which has made it very difficult to quantify and model the sprawl [22].

Most methods for identification of urban sprawl can be categorized as relative and absolute methods [23]. Relative measures compare urban sprawl indicators among different times usually with data of population density or satellite map [23–25], which is often limited by the resolution of data and

faces problems of the definition and recognition of urban built-up areas. Relative measurements allow analysts to conclude whether the study area is sprawling or not, while absolute assessments clearly distinguish compact cities from sprawled ones [26], with spatial metrics of landscape ecology, multiple statistic indices such as employment, resource consumption, population, living quality and architecture aesthetics. Similarly, the lack of a threshold value between harmful sprawl and benign expansion still challenge the outcomes of such methods. Contrary to the uncertainty in such a threshold value, the entropy method presents a clear value of urban sprawl measurement on a general scale [27–31], due to the entropy being able to measure the uneven degree of particle distribution in the system. Bhatta argues that the entropy method is more spatial, robust and static than others [20,23]. However, the entropy method cannot provide a spatial result that indicates which area is sprawling. This article hereby argues that the combination of the entropy method and urban sprawl indices presents an ideal way of identifying urban sprawl.

As Ewing claims, they “know it when they see it” [32] or can “identify urban sprawl based on their judgment” [33]—a simple and universally accepted sprawl indicator deserves exploration, one that could be implemented for a wide range of locations and explained straightforwardly to decision makers [34]. Although the driving forces underlying urban sprawl differentiate throughout the world [35,36], there still exists well-accepted common sense on spatial features of urban sprawl: leapfrog development, low-density and unlimited outward expansion [8]. Phelps and Silva indicate that urban interstices at four geographic scales will influence the measure of urban sprawl, wherein the so-called empty lands are not inert at all [37]. Desalvo and Qing argue that the undeveloped land around an average dwelling could be the main determinant of sprawl, based on the testing of the traditional monocentric model in which population density plays a dominant role [38].

Based on the reinterpretation of urban spatial structure models of the Chicago School with gridded population density map, this paper concludes that there exist five combination patterns of spatial population density in Moore neighborhood units, which helps to identify whether the central unit is in a sprawling area. This paper hereby presents a simplified urban sprawl measuring method based on K-means clustering with merely the population density of gridded units and their local spatial entropy which calculates the uneven distribution degree of population density of surrounding units. With testing of the case study in Wuhan, located in China, the result clearly shows the urban expansion pattern of Wuhan with the identification of the sprawling area.

The paper is organized as follows: the first section reviews research on methods of measuring urban sprawl. Section 2 provides the methodology for identifying urban expansion with the population density map and FCA local spatial entropy, followed by the introduction of the case study area and source data in Section 3. Section 4 presents the findings on urban sprawl analysis, further discussion is in Section 5. The final section summarizes the main findings and reflects on the methodology.

2. Methods

2.1. Urban Expansion Diagrams of Gridded Population

Batty indicated that urban modeling and planning methods had been evolving with the development of cities [39], while a monocentric city grew into a multicentric city, connected with regional cities, weaving a more complicated system in the global socioeconomic networks. Classical urban spatial structure diagrams were first revealed by the Thünen ring model in Isolated State theory and then by Ernest Burgess ‘s Concentric zone model [40]. More importantly, Alonso’s Bid Rent curve theory perfectly explained the monocentric city model with the distance decay pattern of population density away from the city center [41]. As the city keeps expanding outward, the Hoyt’s Sector Model [42] and Harris and Ullman’s Multiple Nuclei Model [43] indicate that the transportation system and emerging subcenters generated sector and multi-nuclei model, which still locally obeyed the distance decay pattern in population density near the expanding core or corridor. According to common sense on urban sprawl, sprawling unit is leapfrog, discontinuous away from the main

city with a population density lower than central area but still relatively higher than its surrounding area. Thus, based on the different pattern of population density distribution, the diagram of the three classical urban spatial structure models by the Chicago School and the fourth pattern as urban sprawl could be concluded (Figure 1).

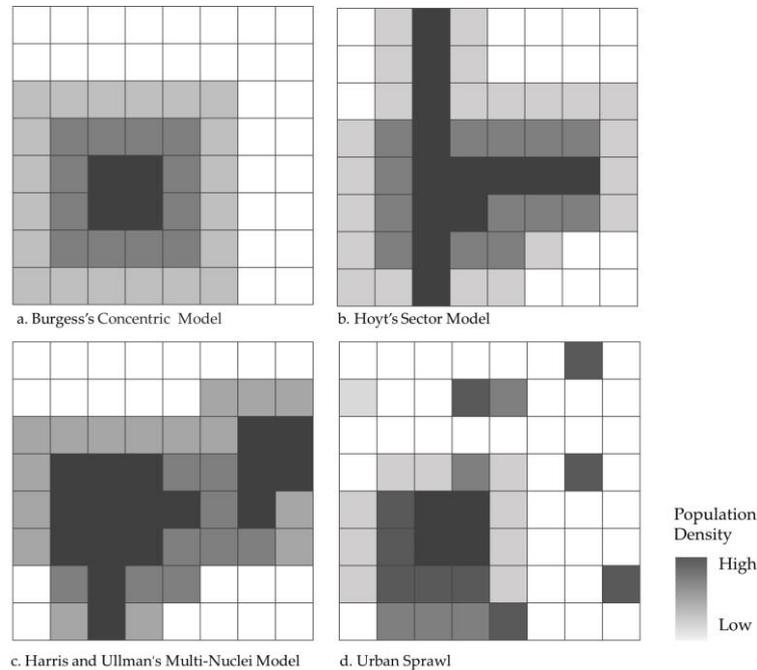


Figure 1. Urban expansion diagrams based on gridded population distribution.

In the previous figure of gridded population density, it is obvious that population density plays the key role in urban expansion analysis, which is consistent with what Ewing and Angel claimed [32,33]. With a Moore neighborhood perspective, this study focused on the land unit and its neighbor units (NU) which could be categorized into five kinds of population density distribution pattern: central area (CA), edge area (EA), rural area (RA), sprawl area (SA) and inner ecological area (IA) (Figure 2). Rather than merely utilizing population density of land unit as a single variable in plenty of related research, this category took the surrounding condition into account.

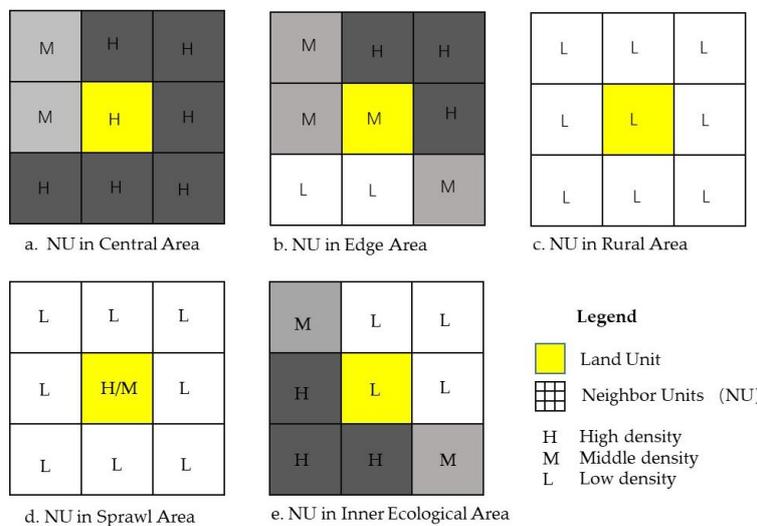


Figure 2. Population Density Distribution Pattern of Land Unit and Neighbor Units.

CA located in the main urban area has relatively higher population density as well as its surrounding units, showing a high evenness in the local neighborhood. Similarly, the RA units possess a same even state but only with lower population density. As SA is defined as a discontinuous unit from the main urban area, it would have a higher population density than its neighbor units, showing an obvious unevenness state in its local area. EA and IA show a middle even state comparing to the uneven state of SA and the even state of CA/ RA, wherein the EA units can be seen as the natural expansion of urban core. This article supposed to categorize all the units by their population density and the different evenness state in terms of local situation, which included the unit and surrounding neighbor units. If grouped into 3, the urban area, rural area and expanding area can be recognized, where the expanding area will contain the edge area, sprawl area. While as grouped into 5, all the above units should be further recognized (Table 1).

Table 1. Groups with Population Density and Local Evenness.

Location	Population Density				
	Land unit	Surrounding units	Local Evenness	Group 3	Group 5
CA	High/Middle	High/Middle	High	1	1
EA	Middle	Middle /Low	Middle	3	2
RA	Low	Low	High	2	3
SA	High/Middle	Low	Low	3	4
IA	Low	High/Middle/Low	Middle	3	5

As the population was already acquired in gridded population map, this experiment only need to calculate the local evenness and define the grouping method, which were defined as FCA local spatial entropy and K-means respectively in this study.

2.2. FCA Local Spatial Entropy

The concept of entropy was first defined in Thermodynamics by Clausius, describing the energy unable to do work during the transferring procedure of energy. Then Boltzmann used entropy as a measurement of disorder in Thermodynamic statistics, which was introduced into information theory and then defined as uncertainty by Shannon [30]. According to the definition of Shannon Entropy, Shannon Entropy can measure the evenness of probability distribution. In a probability set of K (n_1, n_2, \dots, n_k), wherein $\sum n_i = 1$, the evenness of K can be calculated as:

$$H(k) = -\sum P_i \log P_i, \quad (1)$$

$H(k)$ ranges from 0 to $\log K$, while one of the P_i equals 1, $H(k) = 0$; while all the P_i equals $1/k$, $H(k)$ will be the maximum value of $\log K$. In other words, $H(k)$ is related to the amount of K , so when comparing entropy value among data sets of a different amount, the relative entropy is always utilized and written as:

$$R(k) = -\sum P_i \log P_i / \log K, \quad (2)$$

While applying the entropy method for land units to calculate the local evenness, this study presented a method similar to floating catchment area (FCA) method, searching surrounding area within a threshold distance for land unit m , calculating the Shannon Entropy value of the units inside as the evenness value of the central unit m (Figure 3).

Although Batty introduced entropy theory in space-related studies and defined it as Spatial Entropy [27], spatial entropy only can compute the total evenness of population distribution, which had seldom been used locally. While the First Geographic Law of Tobler pointed out, the nearer the lands exist, the more related their attributes are. Because the FCA method with Shannon entropy applied the spatial correlation rule of Tobler's law and is capable to measure the uncertainty locally, this study defined this method as Local Spatial Entropy.

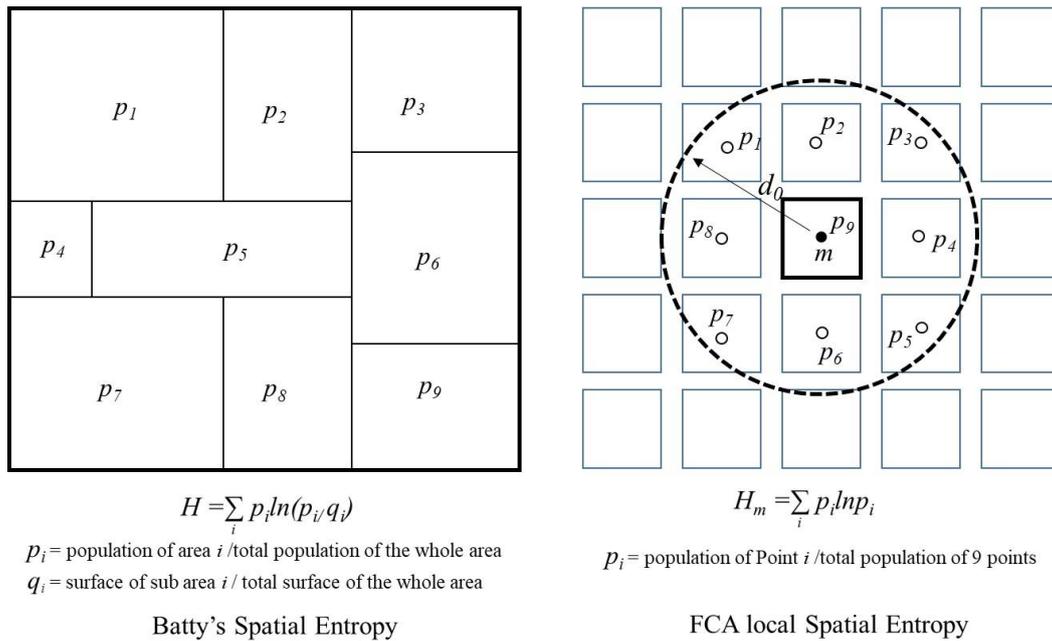


Figure 3. Batty's Spatial Entropy and FCA local spatial Entropy.

The local spatial entropy for gridded units can be calculated in ArcGIS by three steps: (1) Use FEATURE TO POINT to generate centroids for every unit; (2) Use POINT DISTANCE within a threshold searching radius to generate local matrix for each centroid, ensuring merely including the surrounding 8 centroids; (3) Calculate the local spatial entropy for the central point based on Shannon Entropy (Equations (1) and (2)) with the population of all the 9 centroids.

2.3. K-Means Clustering

Since urban sprawl was supposed to be defined by clustering with population density and local spatial entropy, a handy tool suitable for area classification would be preferred. Although there exist many statistical packages, the commonly used K-means clustering was chosen according to Vickers's comparison on cluster analysis methods [44]. K-means clustering is capable to partition units by spatial correlation, aiming to make the differences among the units in a group, overall groups, is minimized [45]. Furthermore, the K-means algorithm is suitable for the situation that the number of clusters has already been designated [46], since the previous assumption has divided the cells into 3 or 5 groups. Besides, the K-means algorithm is distance-based, taking distance as the evaluation index of similarity, that is, the closer the distance between two objects is, the larger the similarity is. The algorithm considers the cluster to be composed of objects that are close together, so the compact and independent cluster is the ultimate target. Considering the division of the data set X into k groups using Euclidean distance, the minimum of Sum of the Squared Error (SSE) is the objective of clustering algorithm:

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} (x_i - C_i)^2 \tag{3}$$

Wherein, k is the number of clusters, x is a set of value in a certain cluster with a mean of C_i .

Admittedly, the uncertainty in defining the number of clusters is the main challenge for K-means clustering, if the number is unknown. Normally, the higher Pseudo F-statistic is, the number would be better, however, when the F-Statistic reaches its max value, the algorithm can result in a local optimum rather than a global one. That is to say, with k increasing, the sample partitioning will be more refined, SSE will definitely become smaller. Wherein, when k is smaller than an ideal number of clusters, the SSE will decrease largely, while k is larger than that, SSE will decrease slowly. Such pattern constructs an elbow shape of SSE and k , or F—Statistic and k (Figure 4), which is also called

the Elbow method [47]. It is an ideal choice to select the k value at elbow point or slightly larger than that, rather than at the maximum value while using F-statistic.

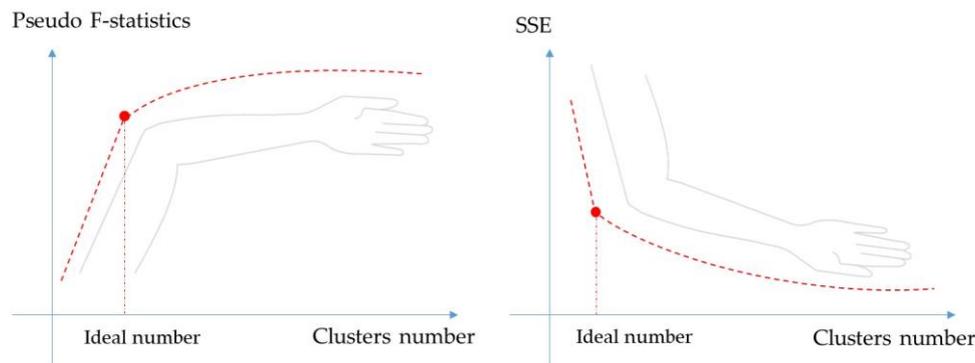


Figure 4. Elbow Method for determining the number of clusters.

K-means clustering can be performed in ArcGIS with a Pseudo F-statistics report, wherein the Group Analysis in toolbox has a built-in module of K-means algorithm while setting NO SPATIAL CONSTRAINT for the Spatial Constraints parameter, which makes the process more convenient.

3. Data

3.1. Study Area

The study area is the Wuhan city, located in central China (Figure 5). As the capital of Hubei province and one of the nine National Central Cities of China, Wuhan is the most populous city in Central China. The city boasts abundant mountain and water resources and is divided into the “Three Towns” of Wuchang, Hankow and Hanyang by the Yangtze River and Han river. Limited by the complicated geographic condition, Wuhan keeps expanding along the Yangtze River and the main inter-provincial roads, especially the north mountain area and south lake area extremely affect the expansion. As the same as other Chinese cities [35,36], the urban expansion of Wuhan has been largely driven by government, making more enclaves outside the urban core.

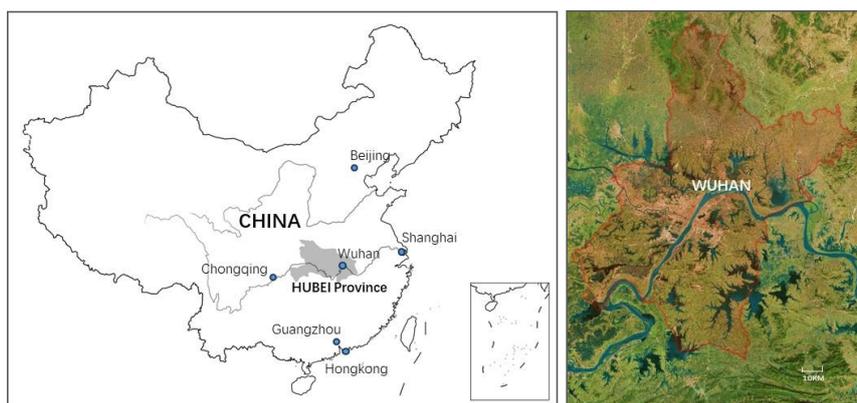


Figure 5. Location of Wuhan, Hubei province in China.

3.2. Open Gridded Population Data of GeodataCn

The gridded population datasets in 2010 was obtained from National Earth System Science Data Sharing Infrastructure of China (<http://www.geodata.cn>), which covered mainland China with a spatial resolution of 1 km, with land use regression model based on census data at the county level and land use data at a scale of 1:100,000 in national scale. Land use data contained urban and rural built-up

area, grassland, forest land and etc., in which built up area made an important role in disaggregating population, hence the GeodataCn represented a static combining population density of both work-time and non-worktime. The GeodataCn had been tested having an estimation accuracy of above 50% in a general national scope [48], which may result in uncertainty in a city level.

3.3. Gridded Population Generated by Mobile Phone Data

To further verify the assumption, this study took a gridded population data generated by mobile phone data. Data of phone call records used in the present study was provided by a partner telecommunication operator whose market share was about 60%, verified for representing whole population distribution proportionally in Wuhan [49]. Mobile phone data of 7,300,000 users in November 2015 in Wuhan City was used in the study. Data was pre-processed, eliminating all privacy-related information. With a similar land use regression model and two-step floating catchment area method (2SFCA), the population of work time was disaggregated to gridded land units, in which the accuracy of the gridded population was verified with an adjusted R^2 of about 80% [50]. The mobile phone data in 2015 provided a verification of the urban sprawl pattern of 2010 and a comparison of the different accuracy of data source (Figure 6). Theoretically, the work time population was more capable to help identify the structure of urban expansion, however, the residential area with few populations in the daytime may influence the result of identifying urban sprawl.

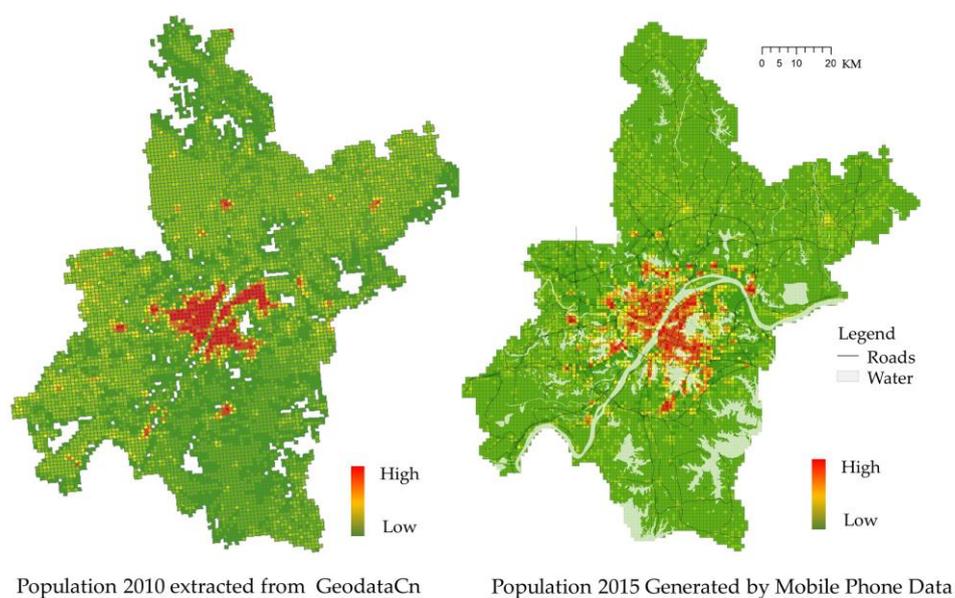


Figure 6. Gridded Population Density Map.

4. Results

Extracted from gridded population map of GeodataCn, the population density map of Wuhan could be easily symbolized, the white area inside the map were units with no population, most of which existed river, lakes and mountains. With the calculation of local spatial entropy base on population, local evenness of grid units could be mapped. However, the units along the border of Wuhan showed obviously lower value than nearby units, because the border units have fewer neighbor units, make the k and $\log K$ smaller. Thus, the relative entropy was further calculated to diminish the influence of k , the number of nearby units. It can be seen in the result mapping that the relative entropy shows a more continuous distribution of evenness value than original Shannon Entropy, especially, the border effect has been eliminated (Figure 7). While applying Group Analysis in ArcGIS with a value of population and local spatial entropy, original population density map was categorized into three groups and five groups (Figure 8) with summary reports (Tables 2 and 3).

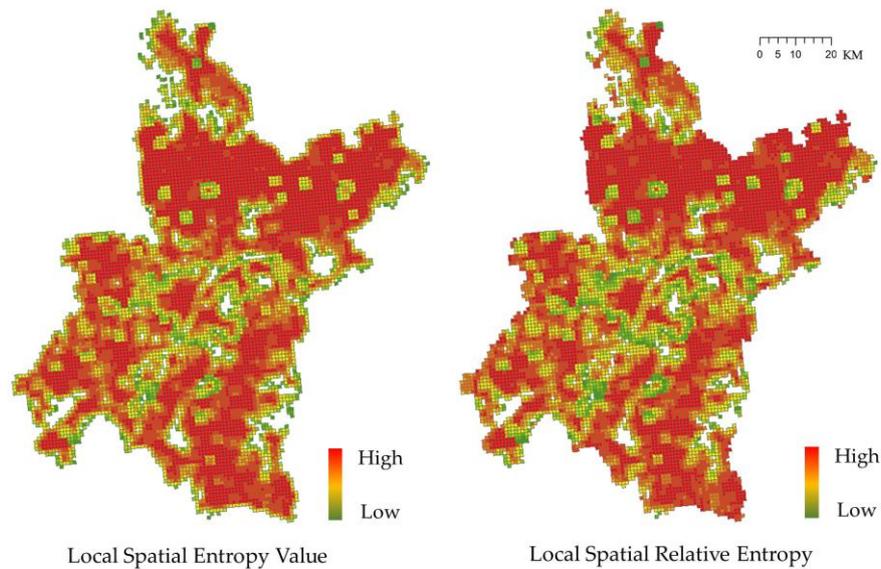


Figure 7. Calculation of Local Spatial Entropy.

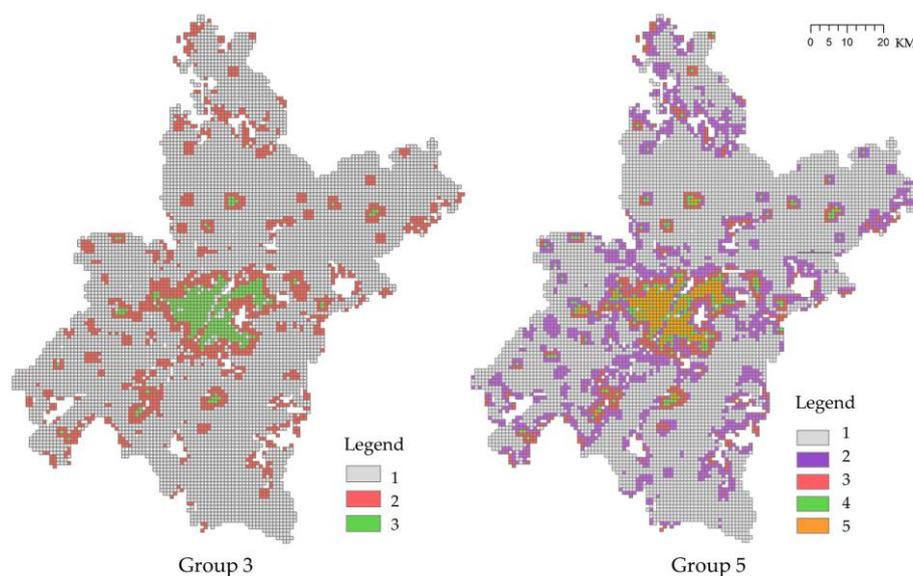


Figure 8. K-means Clustering in Group 3 and Group 5 of Population 2010.

Table 2 shows that in 3-group clustering, the green units in the left picture of Figure 6 share relative high mean population density of 14,278 people/km² and high relative entropy of 0.94, reflecting the feature of Central Area. The gray units with a lower mean population of 449 people/km² and relative entropy of 0.96 also reflect the feature of Rural Area. Obviously, the red ones are the expanding area and it is not difficult to distinguish the urban sprawl area and edge area from whether it is continuous from the central area, which is aligned to Ewing's claim of "knowing it by seeing it."

Table 3 shows that in 5-group clustering, the green and yellow units in the right picture of Figure 6 reflect features of central area with high population and entropy value, which could be recognized as urban area and suburban area. The other ones share lower population but with different entropy value, wherein the red ones reflect an expanding pattern of edge area, the gray ones reflect the pattern of rural area, the purple ones could be supposed to be in the sprawling area.

Table 2. Variable-Wise Summary of Group 3 with K-means Algorithm.

Population $R^2 = 0.88$					
Group	Mean	Std. Dev.	Min	Max	Share
1	449.6209	412.15	1	6529	0.3741
2	764.2076	1335.67	1	10,834	0.6207
3	12,738.2156	2947.59	6529	17,453	0.6259
Total	976.4951	2515.30	1	17,453	1.0000
Relative Entropy $R^2 = 0.68$					
Group	Mean	Std. Dev.	Min	Max	Share
1	0.9542	0.0416	0.8407	1.0000	0.2155
2	0.7293	0.1009	0.2610	0.8521	0.7999
3	0.9125	0.0780	0.5963	0.9999	0.9999
Total	0.9141	0.1022	0.2610	1.0000	1.0000

Table 3. Variable-Wise Summary of Group 5 with K-means Algorithm.

Population $R^2 = 0.88$					
Group	Mean	Std. Dev.	Min	Max	Share
1	449.7935	249.65	1	4087	0.2341
2	402.3769	567.8729	1	3693	0.2116
3	501.9186	760.2594	1	5689	0.3259
4	6880.2183	2224.644	3550	12,781	0.5289
5	14,278.1528	1757.8904	9700	17,453	0.4442
Total	976.4951	2515.3013	1	17,453	1
Relative Entropy $R^2 = 0.68$					
Group	Mean	Std. Dev.	Min	Max	Share
1	0.9693	0.0251	0.9073	1.0000	0.1254
2	0.8455	0.0456	0.7414	0.9075	0.2248
3	0.6366	0.0885	0.261	0.7405	0.6489
4	0.8089	0.0832	0.3401	0.9913	0.8813
5	0.9454	0.0492	0.7684	0.9999	0.3133
Total	0.9141	0.1022	0.261	1.0000	1.0000

Although Figure 9 indicates that the maximum of group number is 13 while the Pseudo F-Statistic gets the maximum value, according to the Elbow method, the 5-group and 3-group clustering are preferred because 3 is the ideal k value at the elbow point. Meanwhile, 5-group also shows a relative high F-statistic value and R^2 with an overlapped value of media and mean. In conclusion, the 3-group clustering can indicate the expanding area, the 5-group with a higher R^2 of 0.94 and 0.84 can further point out the units in urban sprawl area.

Comparing the grouping result map of Figure 8 and the right image of Figure 6, in other words, the expanding trend recognized in 2010 and the existing population distribution in 2015, it can be concluded that the expanding area near the main core has transferred into central area, while those located far away from the center still keeping undeveloped. The comparison confirms that the methodology can be used to identify the pattern of urban sprawl.

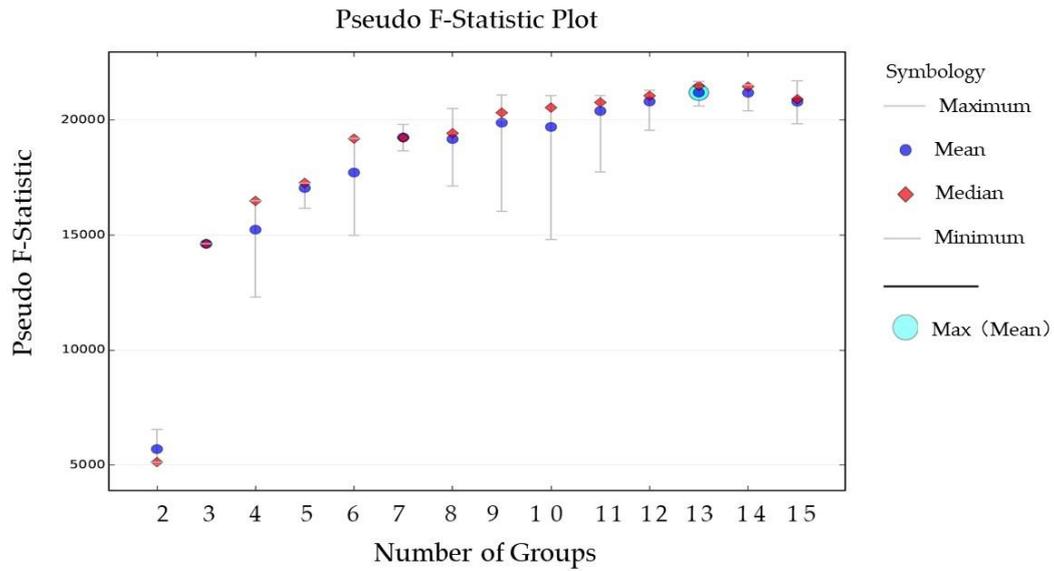


Figure 9. Pseudo F-Statistic Plot of Different Groups.

This study applied the same method to the gridded population 2015 of Wuhan generated by mobile phone data, clustering the land units into 3 and 5 groups (Figure 10) with output summary tables (Tables 4 and 5). Tables 4 and 5 show the same pattern as Tables 2 and 3, as the group 1 is located in the rural area, group 4 and 5 represents the central area. The red units in Group 3 indicate the expanding area, while in 5-Group the red units and purple units show the expanding pattern. Interestingly, Figure 11 indicates that the ideal group number is 3, the same as the previous result for the Geodatacn population.

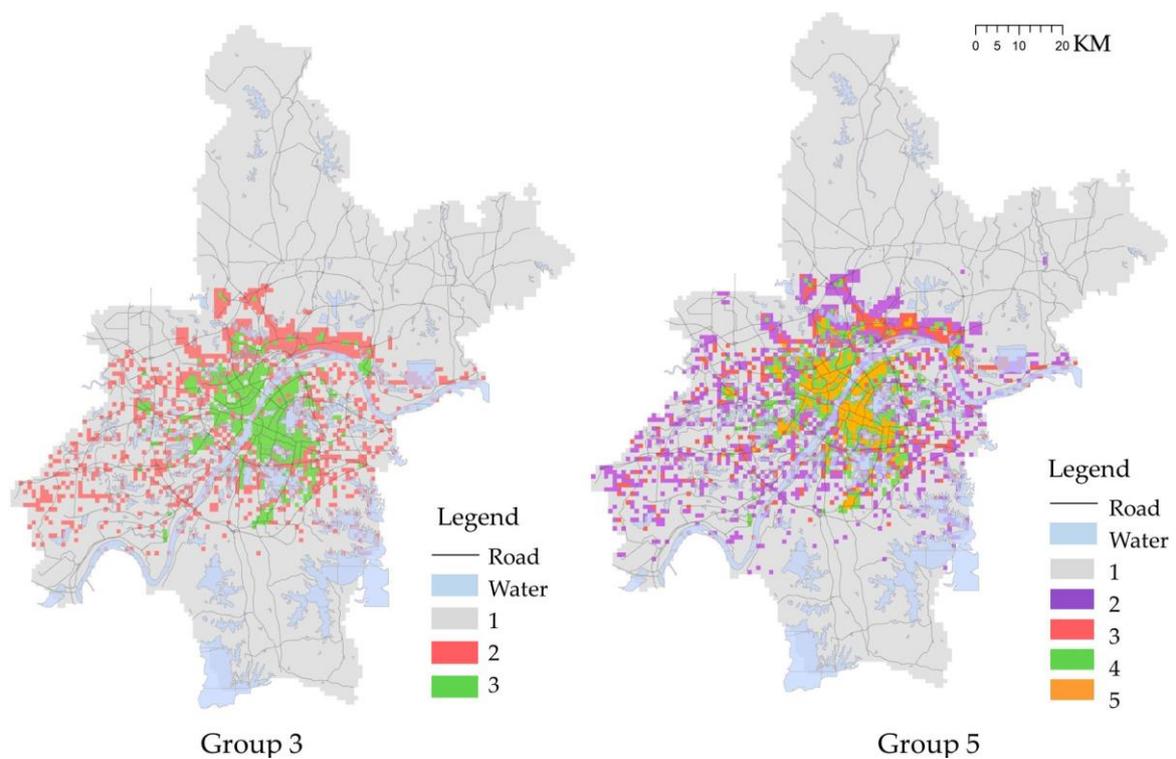


Figure 10. K-means Clustering in Group 3 and Group 5 of population 2015.

Table 4. Variable-Wise Summary of Group 3 with K-means Algorithm.

Population R ² = 0.88					
Group	Mean	Std. Dev.	Min	Max	Share
1	599.6339	573.3265	1	6218	0.1901
2	1129.0176	1466.5203	1	7529	0.2302
3	11,069.4944	3536.1484	5664	32,708	0.8268
Total	1190.7837	2524.6185	1	32,708	1
Relative Entropy R ² = 0.68					
Group	Mean	Std. Dev.	Min	Max	Share
1	0.9748	0.0336	0.8309	1	0.1706
2	0.6904	0.1209	0.0091	0.8463	0.8449
3	0.8277	0.1294	0.3063	0.9973	0.6973
Total	0.9336	0.112	0.0091	1	1

Table 5. Variable-Wise Summary of Group 5 with K-means Algorithm.

Population R ² = 0.88					
Group	Mean	Std. Dev.	Min	Max	Share
1	558.6441	397.6774	1	4578	0.1399
2	757.1577	821.7557	1	4048	0.1238
3	1345.1096	1883.1429	1	10,416	0.3184
4	7135.8354	1867.0962	3752	10,466	0.2053
5	13,846.3772	2696.6543	10,480	32,708	0.6796
Total	1190.7837	2524.6185	1	32,708	1
Relative Entropy R ² = 0.68					
Group	Mean	Std. Dev.	Min	Max	Share
1	0.9809	0.0217	0.8881	1	0.1129
2	0.7968	0.0611	0.6739	0.8917	0.2198
3	0.5585	0.1057	0.0091	0.6902	0.6873
4	0.8447	0.1045	0.5504	0.9983	0.452
5	0.8278	0.1335	0.3063	0.9973	0.6973
Total	0.9336	0.112	0.0091	1	1

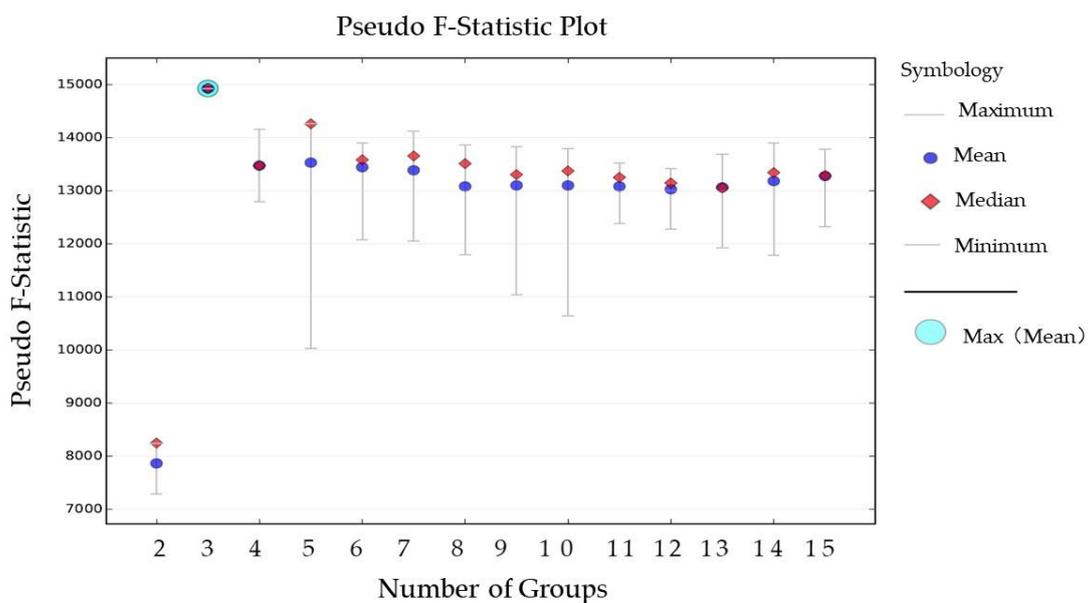


Figure 11. Pseudo F-Statistic Plot of Different Groups.

To further illustrate the difference and similarity in these two data, a four-quadrant diagram was provided with the mean value of population and local spatial entropy in all the groups (Figure 12). As the clustering number increases from 3 to 5, both kinds of data show three features: relatively stable rural units with high entropy value, subdivision of urban units and expanding units. Derived from the separating of urban units, the suburban units are refined with a relatively large amount of mean population. Meanwhile, the expanding units are divided into two groups, one with lower local spatial entropy and larger population, the other with higher entropy and a smaller population, both could be identified as sprawling units in term of different definitions of urban sprawl. In this article, the scattered units are defined as sprawling ones, thus both the east and west side of Wuhan show an obvious sprawling trend, wherein most sprawling units are located along the main roads. Instead, the north side in Wuhan shows a classic Hoyt’s sector pattern, a strip or belt expansion.

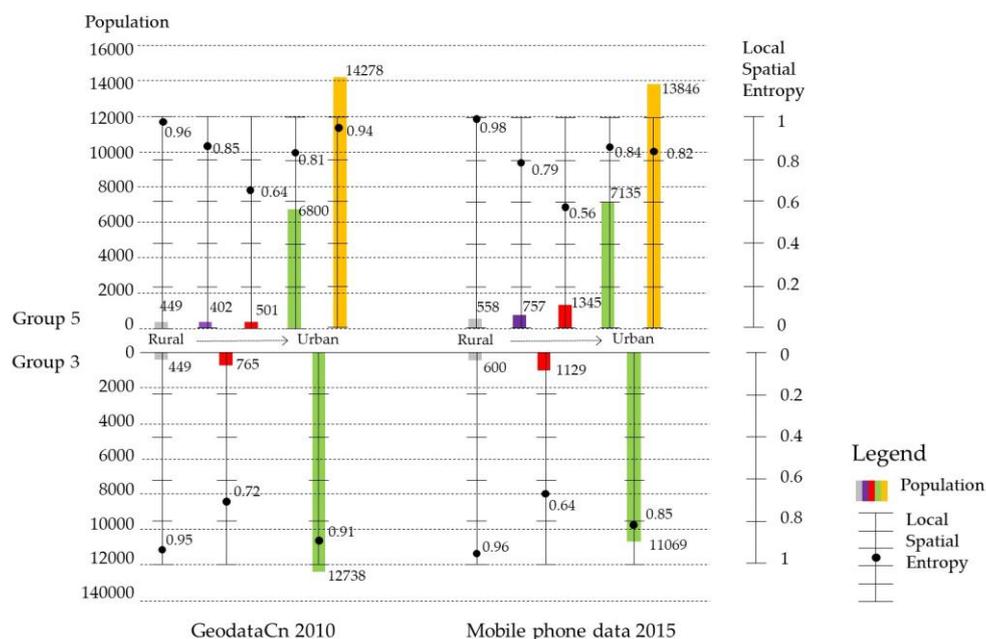


Figure 12. Mean value of Population and Local Spatial Entropy.

A detailed expansion map of Wuhan provided by local planning bureau with the built-up area information in 2010 and 2015 (Figure 13) was compared to further testify the result of the proposed method. Not only most of the strip-growth pattern in the north of Wuhan but also the expansion area in east and south side is overlapped, which show the applicability of the proposed fast identification method. Moreover, the scatter sprawling units detected by this article shows more information about the urban sprawling pattern of Wuhan. Generally, Wuhan is expanding along the Yangtze River, keeping the restraints on the development in the direction of the south and north, where exist ecological barriers of mountains and lakes. As the east side of Wuhan has been planned as a new city with the high-tech industrial park, the sprawling state can be seen as a result of upside down driving force. However, it is noteworthy that the massive sprawling area also emerges in the west direction, where is located in the plains and near Hanyang, which is a city with a famous automobile industrial park.

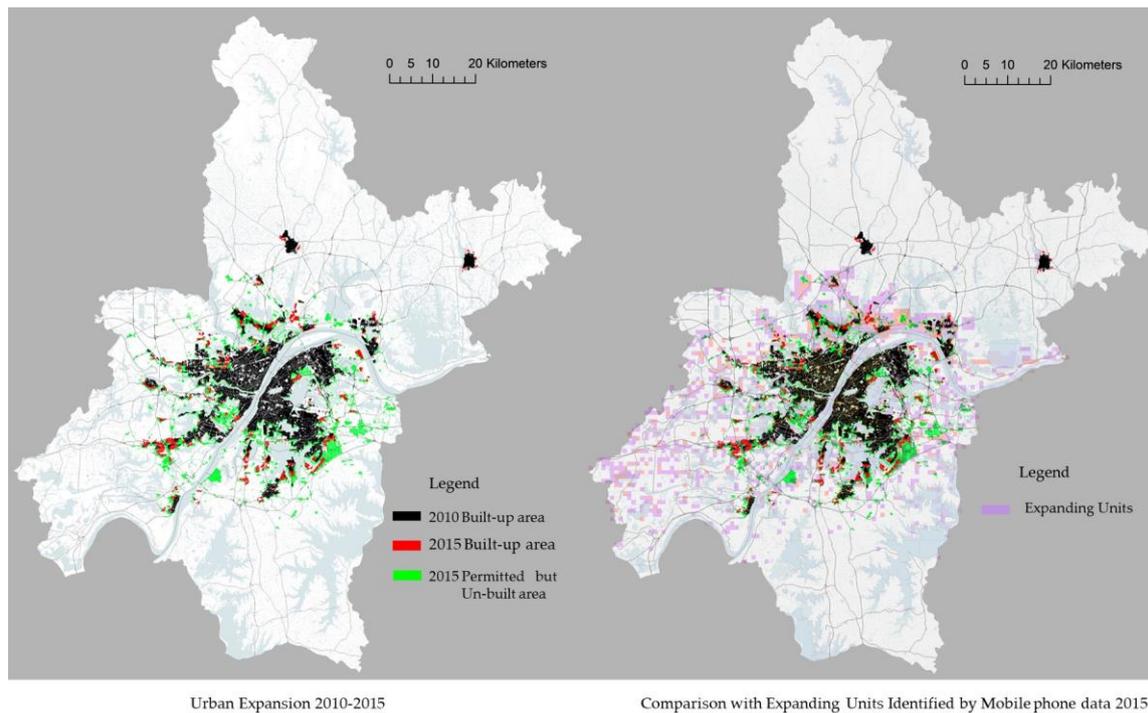


Figure 13. Comparison with Built-up area changes.

5. Discussion

The research we have done suggests that gridded population can be the only source data to identify the pattern of urban sprawl with K-means clustering based on population and local spatial entropy. These findings are understandable because grid units located in different location have the corresponding level of population and unevenness of surrounding situation, wherein urban sprawl area shows a relatively high population and low local spatial entropy, different from other units, making it easy to be identified.

The ideal cluster number of 3 at the Pseudo F-statistics elbow point in both GeodataCn and mobile phone data further verifies that K-means clustering is suitable to identify the expanding land units from the urban and rural area. On the other hand, the coincidence that Pseudo F-statistics reaches its max value at the elbow point in Figure 9 indicates that the higher resolution population data provides stronger support for the method.

These results agree with Ewing [32] and Angel [33]'s statement, urban sprawl pattern can be recognized by seeing it, moreover, the two variables of population and local spatial entropy support what Joseph [38] and Burchfield [51] have pointed out that the determinant of urban sprawl are population and the amount of undeveloped land around an average dwelling. However, the procedure and data utilized in this research are more simple and straightforward than previous studies. This approach corresponds to the classical urban spatial structure theory of the Chicago School and is highly consistent with Alonso-Muth-Mills' work, which explained the importance of population density in identifying urban structures [52]. On the other hand, the interstices in the scattered sprawling units verify the pending feature of undeveloped land in urban expansion [37], which should be further explored.

Based on the comparison of output result by open population data and mobile phone data, the finding of this study is restricted to the resolution and accuracy of the data source. Moreover, the FCA local spatial entropy method applied in this research only considers the surrounding 8 units, further study of which threshold distance should be selected is needed.

6. Conclusions

This study suggests the identification of urban sprawl based on K-means clustering with the merely gridded population and its local spatial entropy, wherein the Elbow method is used to verify the group number. Based on the result of the experiments, such an assumption is verified, which further indicates that the role of population density deserves more exploration. The other important contribution of this study is the application of FCA local spatial entropy, which combines Shannon entropy, Tobler's first law of geography and Moore neighborhood, focusing on the measurement of local spatial unevenness.

The present study focuses on the measurement based on human activities, other than physical environment emphasized by most remote sensing method, which may result in a malfunction in identifying the typical sprawling pattern of Ghost towns [53], where few populations live in newly developed cities. Another limitation is that the accuracy of population data which may influence the discrimination of urban sprawl units, however, with the quick development of urban data environment, population data will be more precise, real-time, thus the method of fast identification for urban sprawling would help government and planners makes quicker and more flexible decisions. In prospect studies, comparison among cities or sprawling areas within the same city-region could be conducted, exploring the driving force underlying urban sprawl.

Author Contributions: L.L. and H.W. conceived and designed the experiments; L.L. performed the experiments; Z.P. and J.Z. acquired and analyzed the data; Y.Y. and H.J. contributed reagents/materials/analysis tools; H.W. and L.L. wrote the paper.

Funding: The research is funded by China Postdoctoral Science Foundation (No. 2016M600609); National Science Fund for Young Scholars (No. 51708425); and China Postdoctoral Science Foundation (No. 2016M602357).

Acknowledgments: The open population data is supported by "National Earth System Science Data Sharing Infrastructure, National Science & Technology Infrastructure of China (<http://www.geodata.cn>)".

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

References

1. Brueckner, J.K. Urban Sprawl: Diagnosis and Remedies. *Int. Reg. Sci. Rev.* **2000**, *23*, 160–171. [[CrossRef](#)]
2. Vietz, J.G.; Rutherford, I.D.; Walsh, C.J.; Chee, Y.E.; Hatt, B.E. The Unaccounted Costs of Conventional Urban Development: Protecting Stream Systems in an Age of Urban Sprawl. In Proceedings of the Australian Stream Management Conference, Townsville, QLD, Australia, 31 July 2014.
3. Heckman, C.J. Public Parks and Shady Areas in Times of Climate Change, Urban Sprawl, and Obesity. *Am. J. Public Health* **2017**, *107*, 1856–1858. [[CrossRef](#)] [[PubMed](#)]
4. Frumkin, H. Urban Sprawl and Public Health. *Public Health Rep.* **2002**, *117*, 201. [[CrossRef](#)]
5. Wu, F.; Xu, J.; Yeh, A.G. *Urban Development in Post-Reform China: State, Market, and Space*; Routledge: Abingdon, UK, 2006.
6. Christiansen, P.; Loftsgarden, T. Drivers Behind Urban Sprawl in Europe. *TØI Rep.* **2011**, *1136*, 2011.
7. Sturm, R.; Cohen, D.A. Suburban Sprawl and Physical and Mental Health. *Public Health* **2004**, *118*, 488–496. [[CrossRef](#)] [[PubMed](#)]
8. Burchell, W.R.; Mukherji, S. Conventional Development Versus Managed Growth: The Costs of Sprawl. *Am. J. Public Health* **2003**, *93*, 1534–1540. [[CrossRef](#)] [[PubMed](#)]
9. Alberti, M. The Effects of Urban Patterns on Ecosystem Function. *Int. Reg. Sci. Rev.* **2005**, *28*, 168–192. [[CrossRef](#)]
10. Zhang, Y.; Yang, Z.; Li, W. Analyses of Urban Ecosystem Based on Information Entropy. *Ecol. Model.* **2006**, *197*, 1–12. [[CrossRef](#)]
11. Wang, H.; Ning, X.; Zhu, W.; Li, F. Comprehensive Evaluation of Urban Sprawl on Ecological Environment Using Multi-Source Data: A Case Study of Beijing. ISPRS—International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2015, XLI-B8, 1073–77. Available online: <https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLI-B8/1073/2016/isprs-archives-XLI-B8-1073-2016.pdf> (accessed on 29 July 2018).

12. Encarnaç o, S.; Gaudiano, M.; Santos, F.C.; Tened rio, J.A.; Pacheco, J.M. Urban Dynamics, Fractals and Generalized Entropy. *Entropy* **2013**, *15*, 2679–2697. [[CrossRef](#)]
13. Sullivan, C.W.; Lovell, S.T. Improving the Visual Quality of Commercial Development at the Rural–Urban Fringe. *Land. Urban Plan.* **2006**, *77*, 152–166. [[CrossRef](#)]
14. Burchell, W.R.; Shad, N.A.; Listokin, D.; Phillips, H.; Downs, A.; Seskin, S.; Davis, J.S.; Moore, T.; Helton, D.; Gall, M. *The Costs of Sprawl-Revisited*; Transportation Research Board: Washington, DC, USA, 1998.
15. Ewing, R.H. Characteristics, Causes, and Effects of Sprawl: A Literature Review. In *Urban Ecology*; Springer: Boston, MA, USA, 2008; pp. 519–535.
16. Frenkel, A.; Ashkenazi, M. The Integrated Sprawl Index: Measuring the Urban Landscape in Israel. *Ann. Reg. Sci.* **2008**, *42*, 99–121. [[CrossRef](#)]
17. Knaap, G.; Talen, E.; Olshansky, R.; Forrest, C. *Government Policy and Urban Sprawl*; Illinois Department of Natural Resources, Office of Realty and Environmental Planning: Springfield, IL, USA, 2000.
18. Tsai, Y.-H. Quantifying Urban Form: Compactness Versus ‘Sprawl’. *Urban Stud.* **2005**, *42*, 141–161. [[CrossRef](#)]
19. Yue, W.; Zhang, L.; Liu, Y. Measuring Sprawl in Large Chinese Cities Along the Yangtze River Via Combined Single and Multidimensional Metrics. *Habitat Int.* **2016**, *57*, 43–52. [[CrossRef](#)]
20. Bhatta, B. *Analysis of Urban Growth and Sprawl from Remote Sensing Data, Advances in Geographic Information Science*; Springer: Berlin/Heidelberg, Germany, 2010.
21. Galster, G.; Hanson, R.; Ratcliffe, M.R.; Wolman, H.; Coleman, S.; Freihage, J. Wrestling Sprawl to the Ground: Defining and Measuring an Elusive Concept. *Hous. Policy Debate* **2001**, *12*, 681–717. [[CrossRef](#)]
22. Wilson, H.E.; Hurd, J.D.; Civco, D.L.; Prisloe, M.P.; Arnold, C. Development of a Geospatial Model to Quantify, Describe and Map Urban Growth. *Remote Sens. Environ.* **2003**, *86*, 275–285. [[CrossRef](#)]
23. Bhatta, B.; Saraswati, S.; Bandyopadhyay, D. Quantifying the Degree-of-Freedom, Degree-of-Sprawl, and Degree-of-Goodness of Urban Growth from Remote Sensing Data. *Appl. Geogr.* **2010**, *30*, 96–111. [[CrossRef](#)]
24. Singh, B. Urban Growth Using Shannon’s Entropy: A Case Study of Rohtak City. *Int. J. Adv. Remote Sens. Gis* **2014**, *3*, 544–552.
25. Torrens, P.M. A Toolkit for Measuring Sprawl. *Appl. Spat. Anal. Policy* **2008**, *1*, 5–36. [[CrossRef](#)]
26. Al-Sharif, A.A.A.; Pradhan, B.; Abdullahi, S. *Urban Sprawl Assessment*; Springer International Publishing: Cham, Switzerland, 2017.
27. Batty, M. Spatial Entropy. *Geogr. Anal.* **1974**, *6*, 1–31. [[CrossRef](#)]
28. Batty, M. Entropy in Spatial Aggregation. *Geogr. Anal.* **1976**, *8*, 1–21. [[CrossRef](#)]
29. Batty, M.; Morphet, R.; Masucci, P.; Stanilov, K. Entropy, Complexity, and Spatial Information. *J. Geogr. Syst.* **2014**, *16*, 363–385.
30. Cabral, P.; Augusto, G.; Tewolde, M.; Araya, Y. Entropy in Urban Systems. *Entropy* **2013**, *15*, 5223–5236. [[CrossRef](#)]
31. Yeh, A.G.O. Measurement and Monitoring of Urban Sprawl in a Rapidly Growing Region Using Entropy. *Photogramm. Eng. Remote Sens.* **2001**, *67*, 83–90.
32. Ewing, R.H. Characteristics, Causes, and Effects of Sprawl: A Literature Review. In *Urban Ecology: An International Perspective on the Interaction between Humans and Nature*; Marzluff, J.M., Shulenberger, E., Endlicher, W., Alberti, M., Bradley, G., Ryan, C., ZumBrunnen, C., Simon, U., Eds.; Springer: Boston, MA, USA, 2008; pp. 519–535.
33. Angel, S.; Parent, J.; Civco, D. Urban Sprawl Metrics: An Analysis of Global Urban Expansion Using Gis. In Proceedings of the ASPRS 2007 Annual Conference, Tampa, FL, USA, 7–11 May 2007.
34. Aurambout, P.J.; Barranco, R.; Lavalle, C. Towards a Simpler Characterization of Urban Sprawl across Urban Areas in Europe. *Land* **2018**, *7*, 33. [[CrossRef](#)]
35. Liu, Y.; Fan, P.; Yue, W.; Song, Y. Impacts of Land Finance on Urban Sprawl in China: The Case of Chongqing. *Land Use Policy* **2018**, *72*, 420–432. [[CrossRef](#)]
36. Tian, L.; Li, Y.; Yan, Y.; Wang, B. Measuring Urban Sprawl and Exploring the Role Planning Plays: A Shanghai Case Study. *Land Use Policy* **2017**, *67*, 426–435. [[CrossRef](#)]
37. Phelps, A.N.; Silva, C. Mind the Gaps! A Research Agenda for Urban Interstices. *Urban Stud.* **2018**, *55*. [[CrossRef](#)]
38. Desalvo, S.J.; Su, Q. The Determinants of Urban Sprawl: Theory and Estimation. *Int. J. Urban Sci.* **2018**, 1–17. [[CrossRef](#)]

39. Batty, M. *Fifty Years of Urban Modeling: Macro-Statics to Micro-Dynamics*; Physica-Verlag HD: Heidelberg, Germany, 2008.
40. Burgess, E.W. The Growth of the City: An Introduction to a Research Project. *City* **2008**, *18*, 71–78.
41. Alonso, W. *Location and Land Use*; Harvard University Press: Cambridge, MA, USA, 1964.
42. United States Federal Housing Administration; Hoyt, H. The Structure and Growth of Residential Neighborhoods in American Cities. *Development* **1941**, *19*, 453–454.
43. Harris, D.C.; Ullman, E.L. The Nature of Cities. *Ann. Am. Acad. Political Soc. Sci.* **1945**, *242*, 7–17. [[CrossRef](#)]
44. Dan, V.; Rees, P. Creating the Uk National Statistics 2001 Output Area Classification. *J. R. Stat. Soc.* **2007**, *170*, 379–403.
45. Harris, R.; Sleight, P.; Webber, R. Geodemographics, Gis and Neighbourhood Targeting. *J. Direct Data Digit. Mark. Pract.* **2007**, *8*, 364–368.
46. Everitt, S.B.; Dunn, G.; Everitt, B.S.; Dunn, G. *Cluster Analysis*; Wiley: New York, NY, USA, 2011.
47. Ketchen, J.D.; Shook, C.L. The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. *Strateg. Manag. J.* **1996**, *17*, 441–458. [[CrossRef](#)]
48. Bai, Z.; Wang, J.; Wang, M.; Gao, M.; Sun, J. Accuracy Assessment of Multi-Source Gridded Population Distribution Datasets in China. *Sustainability* **2018**, *10*, 1363. [[CrossRef](#)]
49. Wu, H.; Liu, L.; Yu, Y.; Peng, Z. Evaluation and Planning of Urban Green Space Distribution Based on Mobile Phone Data and Two-Step Floating Catchment Area Method. *Sustainability* **2018**, *10*, 214. [[CrossRef](#)]
50. Liu, L.; Peng, Z.; Wu, H.; Jiao, H.; Yu, Y. Exploring Urban Spatial Feature with Dasymetric Mapping Based on Mobile Phone Data and Lur-2sfcae Method. *Sustainability* **2018**, *10*, 2432. [[CrossRef](#)]
51. Burchfield, M.; Overman, H.G.; Puga, D.; Turner, M.A. Causes of Sprawl: A Portrait from Space. *Q. J. Econ.* **2006**, *121*, 587–633. [[CrossRef](#)]
52. Takahashi, T. Location Competition in an Alonso–Mills–Muth City. *Reg. Sci. Urban Econ.* **2014**, *48*, 82–93. [[CrossRef](#)]
53. Ge, W.; Yang, H.; Zhu, X.; Ma, M.; Yang, Y. Ghost City Extraction and Rate Estimation in China Based on Npp-Viirs Night-Time Light Data. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 219. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).