

Article

A Method for Identifying Urban Functional Zones Based on Landscape Types and Human Activities

Yongcai Jing^{1,2}, Ranhao Sun^{1,2,*}  and Liding Chen^{1,2}

¹ State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China; jingyongcai14@mails.ucas.ac.cn (Y.J.); liding@rcees.ac.cn (L.C.)

² University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: rhsun@rcees.ac.cn

Abstract: The effects of land use and socioeconomic changes on urban landscape patterns and functional zones have been increasingly investigated around the world; however, our knowledge on these effects is still inadequate for sustainably managing urban ecosystems. The urban functional zone (UFZ) refers to a kind of regional space that provides specific functions for human activities and reflects the land use type in a city. They are important for urban planning and exploring urban texture dynamics. UFZs improve understanding of sustainable development for urban ecosystems with extreme environments and unique social backgrounds. However, the identification methods for UFZs are incomplete because of a lack of socioeconomic attributes, as well as their hierarchical relations. Here, we present a hierarchical weighted clustering model to identify UFZs based on the entropy weight method. The data included points of interest (POIs), land use type data, road network data, socioeconomic data, and population density. We found that the adjusted cosine metric and the average criterion were the optimal distance metric and linkage strategy, respectively, to cluster urban zone data. The performance with weighted data was better than that with raw data, and the level of the POI classification scheme and landscape pattern affected the accuracy of identification UFZs. The research indicated that the hierarchical weighted clustering model was a useful method to classify UFZs in order to improve urban planning and environmental management schemes.

Keywords: urban functional zone; hierarchical clustering; points of interest



Citation: Jing, Y.; Sun, R.; Chen, L. A Method for Identifying Urban Functional Zones Based on Landscape Types and Human Activities. *Sustainability* **2022**, *14*, 4130. <https://doi.org/10.3390/su14074130>

Academic Editors: Xiaoping Rui and Runkui Li

Received: 6 March 2022

Accepted: 24 March 2022

Published: 30 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Efforts to make society and its processes more livable necessitate sustainability, which has long been regarded as one of the most significant policy objectives in the world [1]. The spatial patterns of buildings or functional zones affect the urban heat island (UHI) and the sustainability of a city [2]. Traditional techniques for urban functional zone identification, however, are unable to meet the objectives of sustainable urban development because of a lack of hierarchical relations [3]. The “functional zone” is a concept that describes the social and economic properties that satisfy various needs and accommodate diverse human activities in a certain area [4,5]. Urban functional zones share common social and economic activities and are spatially aggregated by diverse geographic objects and semantically abstracted from land uses [3,6]. Urban populations are increasing, with the number of mega-sized cities expected to increase from 10 in 1990 to 41 in 2030 [7]. According to the reports of United Nations, urbanization has been rapid in recent decades, and 68% of the world population are projected to live in urban areas by 2050 [8]. Population density is growing, and the urban area is expanding along with the intensive urban growth [9]. Urban areas directly consume land as their physical footprints expand, resulting in landscape and urban function transformation [10–13]. Strong spatial clustering patterns can also be seen in urban socioeconomic activity [14]. These clustering patterns lead to the generation

of various functional zones to accommodate people's diverse needs for living, working, education, recreation, and public service. UFZs, or basic units for quantitative assessments in urban planning, urban environmental pollution, and other fields, are in great demand. UFZs have an important impact on a city's economy, society, and ecology, and have been increasingly investigated in interdisciplinary studies around the world in relation to factors such as the heat island effect [15–18], runoff characteristics [9,19], urban landscape patterns [20,21], and ecosystem services [22,23].

The UFZ is one of the most effective methods for analyzing urban fabric, conducting urban planning [24,25], and determining the consequences of the urban landscape pattern on various ecological processes [26–28]. However, because making them would require multisource heterogeneous processes and data, urban functional-zone maps are hardly available. Most early studies, on the other hand, were concerned primarily with the identification of pixel-based or object-based image analysis rather than analysis based on urban functions [29–31]. To classify the UFZs, these studies often segmented the study area into regular grids [32,33], administrative divisions [34], or irregular polygons or disjointed blocks by road networks [35,36]. More and more researchers have realized that functional zoning is completely different from land use and land cover classification; thus, functional zones cannot be classified by traditional methods [36–38]. However, fine function zoning is hard work because of the complexity of urban structure and dynamic changes of the function in different periods [39–41]. In recent years, advancements in web mapping, very-high-resolution (VHR) remote sensing, and location-based services (LBS) have provided a more ideal alternative for generating the spatial units of UFZs by recognizing the urban texture and physical properties. Many research studies on mapping UFZs have been conducted by integrating data from multiple sources, such as VHR satellite images and social and human mobility data. Taxi GPS trajectory datasets [36]; cell tower traces [42]; points of interest (POIs) [37,43]; and geotweets, geotagged photos, or check-in data [35,44] are the most commonly used social and human mobility data. Although there have been many improvements due to integrating multisource heterogeneous data, to date, it has still not been possible to characterize functional zones comprehensively and maintain urban sustainable planning, as the hierarchical relations among functional zones have been ignored, which has harmed the representative abilities of UFZ classification methods to date [3,37].

To resolve the issue of accuracy of segmented map and UFZ identification, object-oriented image recognition and topic models were introduced. Object-oriented image recognition takes advantage of the spectral and spatial patterns of geographic objects, and takes spectra, textures, shapes, and the spatial configuration of the landscape into account [3] by using support vector machines (SVM) and artificial neural networks (ANN) [15]. Topic models include k-means clustering [45]; probabilistic latent semantic analysis (PLSA), which was developed by Thomas Hofmann in 1999 [46]; and latent Dirichlet allocation (LDA), which is likely the most widely used topic model today [3]. Nevertheless, one of the critical challenges in this topic is the weighted and hierarchical relations.

Hierarchical clustering is a method of cluster analysis aimed at building a hierarchy of clusters [47]. It gives a nested clustering result in the form of a dendrogram or cluster tree, from which different levels of partitions can be obtained [48]. Because of the rich information produced by applying various strategies and metrics, it has been widely used in many industries. So far, hierarchical agglomerative clustering (HAC) has been the most widely used hierarchical method. HAC is a “bottom-up” approach; each observation starts in its own cluster, and pairs of clusters are merged by moving up the hierarchy. In other words, it starts with clusters each consisting of a single data point and then successively merges the two most similar clusters based on certain similarity metrics [47,48].

In this paper, we propose a general framework integrating land use, POIs, and GDP data to analyze UFZs comprehensively. Comparisons of data combinations, data weighted methods, cluster strategies, and metrics indicators were performed based on HAC. The HAC presented in this study integrated three semantic layers together, i.e., land use

type, population density, and composition of POIs (i.e., human activities), as well as their hierarchical relations. An experiment within the Fifth Ring Road, Beijing, China, was conducted to validate the proposed framework. We aimed to: (1) map UFZs using the presented framework; (2) test the performance of the data combination, data weighted methods, cluster strategies, and metrics indicators; and (3) analyze the spatial pattern of functional zones by using an example in Beijing, China.

2. Materials

2.1. Study Area

Beijing is the capital of China, the world's most populous country. Beijing has become one of the world's fastest expanding cities in recent decades as a result of rapid industrialization and urbanization. The central Beijing area, which is surrounded by multiple ring roads, is made up of several concentric belts of infrastructure and functional zones. The Fifth Ring Road area is the core of the downtown Beijing district, covering 667 km². It is affected by intensified human activities and has a variety of functional zones, such as the educational zones, public zones, recreation areas, and business districts, etc. The study area offers a significant diversity of human activities with distinct urban functional zones. The study area was divided into 336 sub-regions (Figure 1), each with a minimum area of 150,000 m². Segmented zones can be represented by eigenvector consist of amounts or relative amounts of characteristics. Each segmented region is relatively homogeneous in terms of socio-economic function [43].

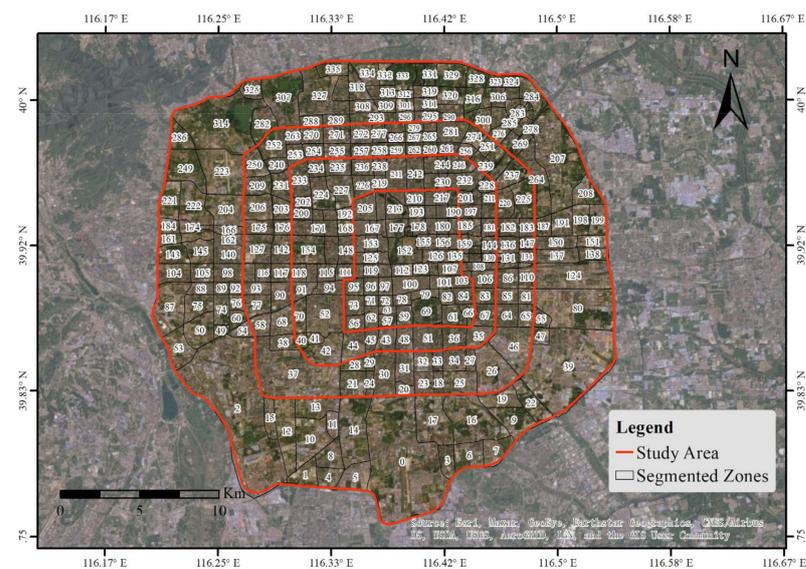


Figure 1. The study area within the fifth-ring Road of Beijing.

2.2. Data Sources

Multi-source data were used in this study, such as POIs, land use type data, road network data, socioeconomic data, and population density. The POIs were obtained through AMap™ (<https://www.amap.com>, accessed 1 September 2018), a web-mapping, navigation, and LBS provider. A total of 572,169 POIs were retrieved in September 2018. Recreation, Catering, Automotive Services, Financial, Education, Public, Health Care Services, Hospitality, Residence, Organizations, and Travel are among the 23 categories of POI data. Furthermore, while 20 of these types are stable categories, the other 3 categories are real-time incidents, such as traffic accidents and road maintenance incidents. For each POI, there are six column properties for each POI: Name, Coordinates, and Categories in three hierarchy levels (composed of primary, secondary, and third-level classes, otherwise called level 1 (L1), level 2 (L2), and level 3 (L3), respectively) (Figure 2). For example, level 1—Education Service, including college, middle school, elementary school, and

kindergarten, can distinguish between the functional properties of the sub-regions. As a result, a data processing framework must be created in order to compute the weight of the comprehensive evaluation of categories at 3 levels, respectively. POIs were used to present the human activities and hierarchical relations. The POIs were divided into 20 primary classes (Table 1), 264 secondary classes, and 868 three-level classes. The land use data [49] with a spatial resolution of 10 m were obtained from Department of Earth System Science/Institute for Global Change Studies Tsinghua University (<http://data.ess.tsinghua.edu.cn/> accessed 1 January 2020). The land-use composition was described by the proportions of urban areas, urban green land, farmland, and woodland. The proportion of various land uses was used to describe land-use heterogeneity. The urban road network data comes from the Open Street Map (OSM) geographic data platform (<https://www.openstreetmap.org/> accessed 1 January 2020). Redundant paths and broken paths were weeded out to represent the functional unites of the study area. The population from WorldPop products of 2017 had a spatial resolution of 1 km × 1km (<https://www.worldpop.org/> accessed 1 January 2020). Statistical socioeconomic data (i.e., population, GDP) in 2017 were obtained from the National Bureau of Statistics.

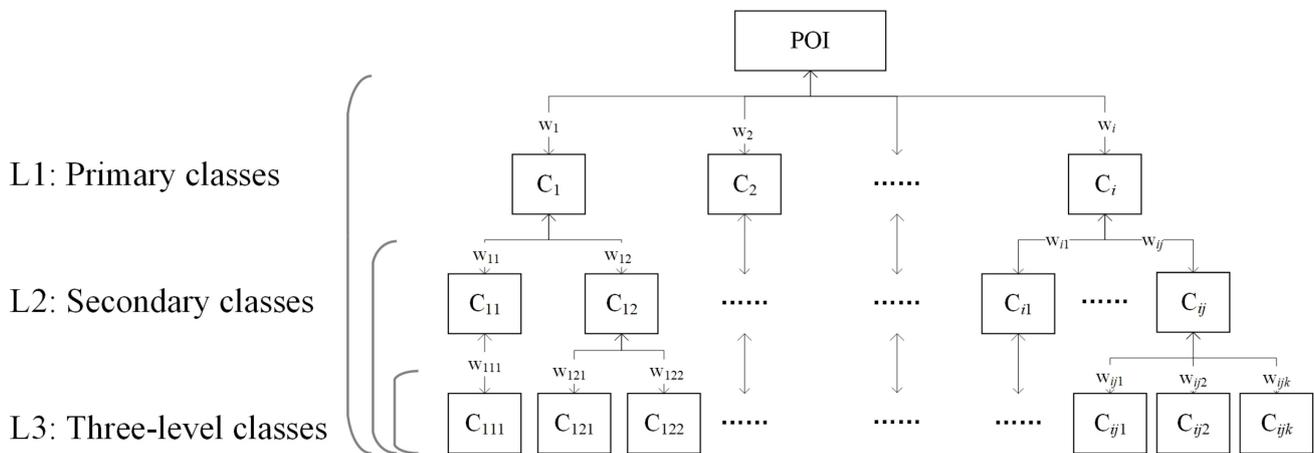


Figure 2. Hierarchical structure of POI properties.

Table 1. Counts and proportion of points of interest types.

ID	Primary Classification	Counts	Proportion
1	Accommodation Service	10,731	1.9%
2	Auto Dealers	780	0.1%
3	Auto Repair	2040	0.4%
4	Auto Service	7293	1.3%
5	Commercial House	21,483	3.7%
6	Daily Life Service	70,141	12.2%
7	Enterprises	66,689	11.6%
8	Finance and Insurance Service	13,965	2.4%
9	Food and Beverages	52,961	9.2%
10	Governmental Organization and Social Group	24,429	4.2%
11	Medical Service	12,574	2.2%
12	Motorcycle Service	321	0.1%
13	Place Name and Address	86,802	15.1%
14	Public Facility	11,331	2.0%
15	Road Furniture	1894	0.3%
16	Science/Culture and Education Service	34,578	6.0%
17	Shopping	95,629	16.6%
18	Sports and Recreation	12,518	2.2%
19	Tourist Attraction	3475	0.6%
20	Transportation Service	46,135	8.0%
	Total	575,769	100.0%

3. Methodology

3.1. The Framework for Identifying UFZs

The segmented regions within the same cluster have similarity characteristic vectors that include the proportion of POIs, land use type, and socio-economic data. The similarity can be gauged by the distance between the two segmented regions. The regions have a high degree of resemblance if the similarity distance is small, and we can expect them to act similarly in terms of urban functions. The larger the distance, the smaller the similarity, indicating that the regions diverge significantly. The characteristic vector of each segmented region can be defined as:

$$R_i = [C_{i,1}, C_{i,2}, \dots, C_{i,n}, Lu_1, \dots, Lu_j, Pop, GDP] \tag{1}$$

Here, R_i is the segmented patch i , and $C_{i,n}$ is the amount of one type in a POI classification scheme at the same level within R_i . $n = 20, n = 264$, and $n = 868$ represents cluster POIs at $L1, L2$, and $L3$, respectively. Lu_j is the proportion of land use type, Pop is the population density, and GDP is the per capital GDP.

As illustrated in Figure 3, the study area was initially segmented into research units by the road network data. The eigenvectors of each segmented research units are then composed of various data combinations based on independence or combination of POIs data at various levels, land use type data, population density, and GDP data. Finally, the Shannon entropy was used to calculate the weight of POI classes at various levels. The results of cluster results by various similarity metric indicators and cluster strategies. Two data processing datasets, three levels of POI classification schemes, six clustering merging strategies with a vector matrix of hierarchical weighted count of POIs within the region, and four similarity distance measure methods are all included in the data.

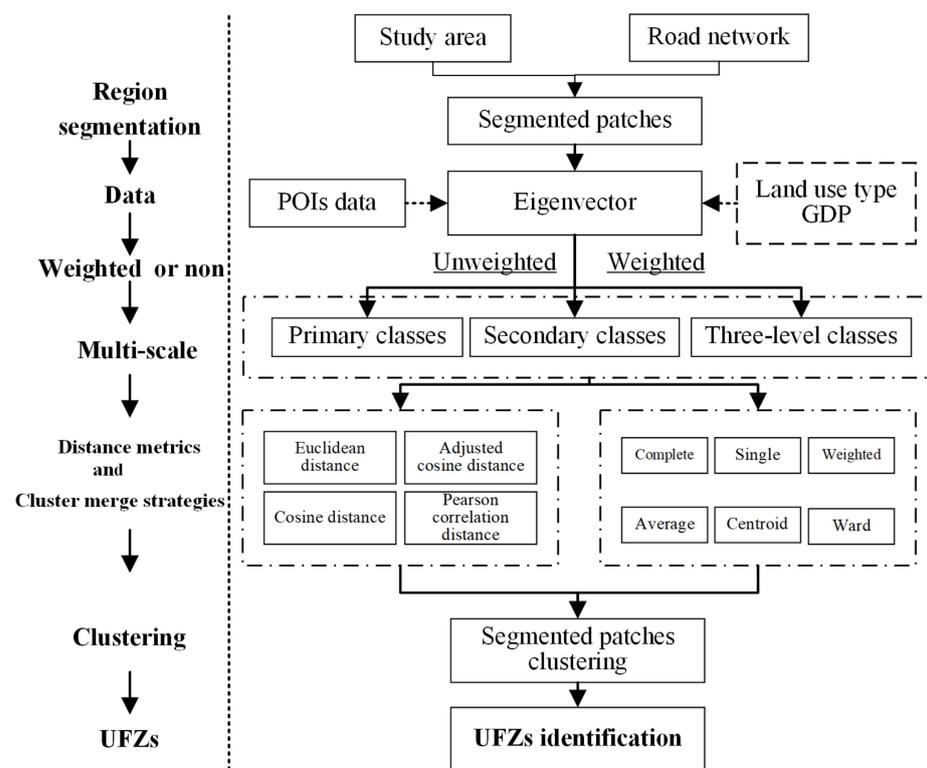


Figure 3. Flowchart of the framework for mapping the UFZs.

3.2. Hierarchical Weighted Clustering Model

Hierarchical agglomerative clustering algorithms represent a popular unsupervised learning technique that seeks to build a hierarchy of clusters and to discover the natural groups of a set of observations. Clustering is the process of grouping samples so that

samples in the same group are as similar as possible, while samples in other groups are as distinct as possible.

For the actual functional label of sub-regions is unknown, we tested different distance measurement methods (Euclidean distance, cosine distance, adjusted cosine distance, and Pearson correlation distance) to categorize functional zones. The Euclidean distance method computes the Euclidean distance between two attribute vectors, which is sensitive to the magnitude of the count of POIs, but not sensitive to the percentage of different features. The cosine distance method computes the cosine distance between two attribute vectors, which is sensitive to the percentage of different features, but not the magnitude of the count of POIs. The adjusted cosine distance computes the cosine distance between two preprocessing attribute vectors by subtracting the mean value. The Pearson by correlation distance method computes the Pearson correlation distance between two attribute vectors.

3.3. Weighting Coefficients and Construct Eigenmatrix

The urban functional zone is influenced by the amount or the weight amount of POIs at each level which characterize the intensity of human activity. We propose the entropy weight method, based on the Shannon entropy theory, integrated with the hierarchical agglomerative clustering method, to balance discrepancy between different subregions, which could contribute to the identification of UFZs as a comparison to compare the clustering results.

The weighting coefficients for different POI types were calculated using the Shannon entropy approach. Shannon entropy is a probability theory-based notion that was developed as a measure of information uncertainty. Since the concept of entropy is well adapted to measuring the relative intensities of contrast criteria [50], it can be used to represent the average intrinsic information transmitted for decision-making. It is a good and practical alternative for us to calculate the weight of POIs type at different levels. On each subtree of the POI classification scheme, we apply the entropy weight method:

Step 1: Standardization of data

Because the data of the metric are not uniform, it is necessary for us to standardize the data. The data were standardized according to the following methods.

$$x'_{Lij} = \frac{x_{Lij} - \min(x_{Lij})}{\max(x_{Lij}) - \min(x_{Lij})} \quad (2)$$

where x_{Lij} is the standardized count of POIs type j within region i on specific scale L , $\min(x_{Lij})$ and $\max(x_{Lij})$ are the minimum and maximum values on a particular POI type j in respective level, respectively. Through the operation, the value are in the range of 0~1.

Step 2: Calculating entropy of information

The entropy of information is a crucial factor to measure the weight of evaluation metric. The high entropy of information indicates that the weight is larger. The following equation shows how to calculate entropy of information:

$$E_{Lj} = -\ln(n)^{-1} \sum_{i=1}^n P_{Lij} \ln P_{Lij} \quad (3)$$

$$P_{Lij} = \frac{x'_{Lij}}{\sum_{i=1}^n x'_{Lij}} \quad (4)$$

where E_{Lj} is entropy of information of each POIs type j a specific scale L , P_{Lij} is the count variance of each POIs type j within region i on a specific scale L , n is the amount of the data on a specific scale L , and x'_{Lij} is the standardized data.

Step 3: Calculation of weight

After calculating the entropy of information, the weight of each metric is determined using the theory of entropy, which indicates the importance of the metric in the evaluation system.

In terms of the weight, the following formula can be used to obtain the weighted value:

$$W_{Lj} = \frac{1 - E_{Lj}}{1 - \sum E_{Lj}}, j = (1, 2, \dots, i) \quad (5)$$

Step 4: Calculation of the weighted value

Following these steps, it is reasonable for us to obtain a comprehensive score of type j of region i on a specific scale L . Therefore, we can evaluate the weighed count of the region i at level L .

$$Z_{Li} = \sum_{j=1}^i x'_{ij} W_{Lj} \quad (6)$$

Step 5: The above steps are repeated with data on other POI classification scheme subtrees and the next top level. The weighted eigenvector matrix on a specific scale can then be obtained.

Step 6: The pairwise similarity distances are calculated for a given pair of nodes, which reflects their distinct degrees.

The HAC algorithm repeatedly identifies the minimal similarity coefficient in the distance matrix to assign the nodes into a linkage tree after constructing the pairwise distances matrix. Updating the pairwise distance matrix is a crucial step, and hierarchical agglomerative clustering can be accomplished using various algorithm [51]. For measuring the distance between the newly formed cluster and original objects, we used five different HAC algorithm methods: single, average, ward, centroid, and complete linkage strategy.

3.4. Evaluation of Clustering Performance

The most common approaches for assessing the quality of clustering results are cophenetic correlation and some internal indices [52]. The cophenetic correlation coefficient compares (correlates) the actual pairwise distances of all samples to those implied by the hierarchical clustering. When the value is closer to 1, the clustering can better preserve the original distances. Suppose that the original dataset x_i is modeled using a cluster method to produce a dendrogram set t_i , the cophenetic correlation coefficient can be denoted as [53]:

$$c = \frac{\sum_{i < j} (x(i, j) - \bar{x})(t(i, j) - \bar{t})}{\sqrt{\left[\sum_{i < j} (x(i, j) - \bar{x})^2 \right] \cdot \left[\sum_{i < j} (t(i, j) - \bar{t})^2 \right]}} \quad (7)$$

where $x(i, j) = |x_i - x_j|$, i.e., the ordinary Euclidean distance between the i th and j th observations. $t(i, j)$ is the dendrogrammatic distance between the model points t_i and t_j . This distance is the height of the node at which these two points are first joined together. We used the cophenetic correlation coefficient to evaluate the performance of all distance metric. Then, we evaluated the performance of cluster results quality of clustering by the Silhouette coefficient, the Calinski–Harabasz index, and the Davies–Bouldin index [54].

4. Results

4.1. The Best Cluster Model Parameters and Strategies

The cophenetic correlation for a cluster tree is defined as the linear correlation coefficient between the cophenetic distances obtained from the tree and the original distances (or dissimilarities) used to construct the tree [55]. As a result, it is a method for detecting the differences among observations in the cluster tree. Table 2 shows HAC with a sample size ($n = 5, n = 20, n = 30$) at different levels. From the cophenetic correlation results, we found that (1) the weighted data processing method performed better than raw data across all distance metrics; (2) the performance of the adjusted cosine distance metric is similar, regardless of whether the weighted or raw data are used; and (3) the optimal clustering merge strategies differs depending on the number and levels of clusters. When cluster number $n = 5$, the cophenetic correlation result was better than other levels using the adjusted cosine distance metric at level 2. When $n = 20$, the cophenetic correlation coefficient result at level 1 was better than other levels, and the adjusted cosine distance metric, synonymous with the “Ward” clustering merge strategies, achieved the max value

of cophenetic correlation coefficient, i.e., 0.909. When $n = 30$, the result at level 1 was better than other levels, and the adjusted cosine distance metric, synonymous with the centroid of clustering merge strategies, achieved the max value of cophenetic correlation coefficient, i.e., 0.929. Overall, the adjusted cosine was the best distance metric, and the performance with weighted data was better than raw data, according to the cophenetic correlation coefficient results.

The dendrogram achieved as a result of clustering process illustrates the number of clusters obtained and their linkage. According to the dendrogram results in Figure 4 and quality curve, as shown in Figure 5, we found that (1) the clustering results with weighted POI and land use data performed better, which indicates that the identification of UFZs should take into account landscape patterns; (2) the optimal combination methods for clustering the UFZs were used for the adjusted cosine distance metrics and the average of the clustering strategy; and (3) the silhouette coefficient was used for the optimal clustering quality metrics, and the optimal number of clusters was 10.

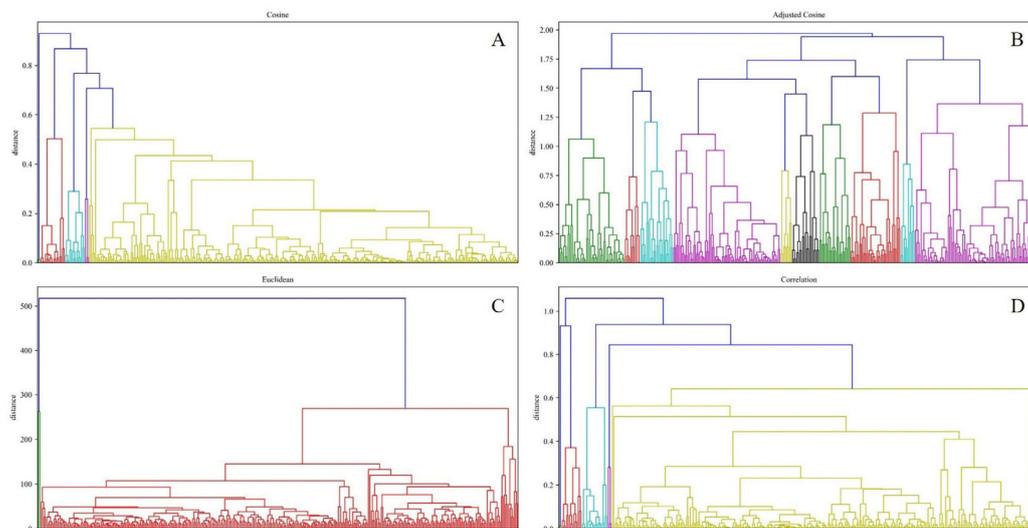


Figure 4. Dendrogram with weighted land use and POIs at level 1. (A) Cosine distance. (B) The adjusted cosine distance. (C) Euclidean distance. (D) Pearson distance.

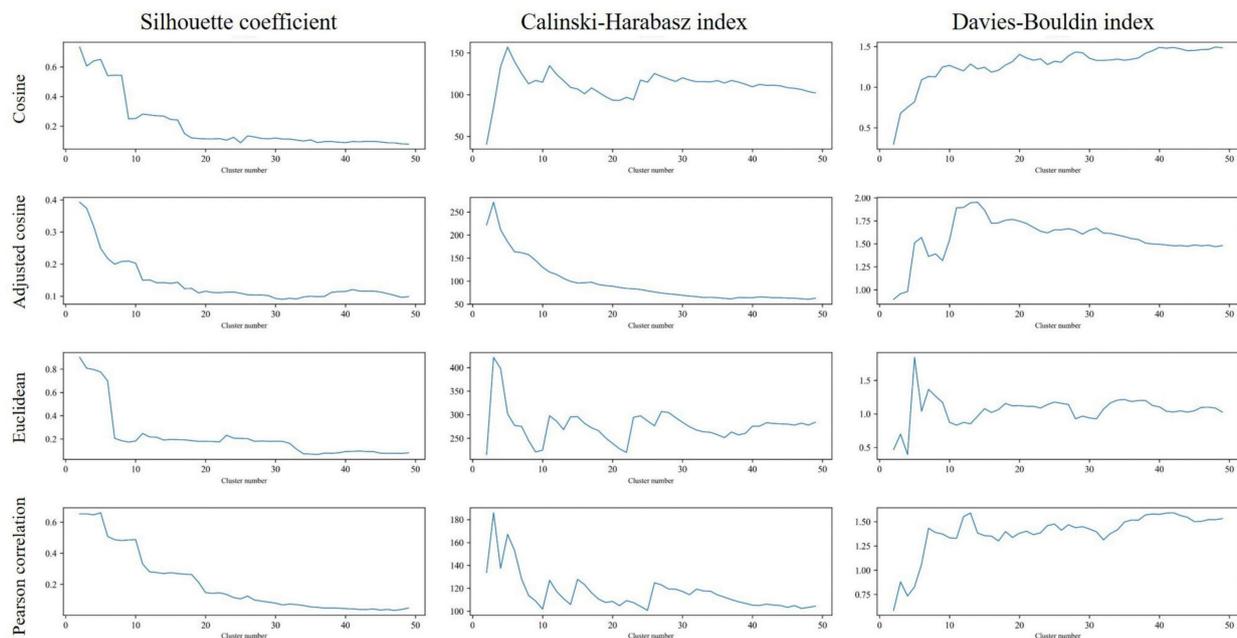


Figure 5. The curve of quality by various distance metrics.

Table 2. The cophenetic correlation coefficient in different method and data set without LULC data.

Distance Metric	Clustering Merge Strategies	The Cophenetic Correlation Coefficient																		
		Weighted POI Data									Raw POI Data									
		n = 5			n = 20			n = 30			n = 5			n = 20			n = 30			
Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
Cosine	Single	0.747	0.924 *	−0.52	0.301	0.919	0.357	0.535	0.764 *	0.367	0.9307 *	0.556	−0.52	0.8218 *	0.708 *	0.357	0.367	0.61	0.367	
	Complete	−0.73	0.86 *	0.831 *	−0.11	0.641	0.368	0.57	0.595	0.445	−0.0946	0.806	0.831	0.33202	0.616	0.368	0.445	0.656	0.445	
	Average	0.699 *	0.913 *	−0.36	0.364	0.773 *	0.549 *	0.679 *	0.705 *	0.536	0.33557	−0.2	−0.36	0.57631	0.55	0.549	0.536	0.658	0.536	
	Centroid	0.757 *	0.913 *	0.831 *	0.364	0.944 *	0.883 *	0.676 *	0.837 *	0.818 *	0.8371 *	0.87 *	0.831 *	0.80752	0.7	0.883	0.818	0.723	0.818	
	Weighted Ward	0.53	0.913 *	0.831 *	0.426	0.96	0.712	0.421	0.718	0.637	0.35358	0.074	0.831	0.53215	0.37	0.712	0.637	0.253	0.637	
		0.304	0.913 *	−0.54	0.642	0.67	0.308	0.549	0.451	0.322	0.27489	0.855 *	−0.54	0.542	0.716	0.308	0.322	0.634	0.322	
The adjusted cosine	Single	0.99 *	0.372	−0.11	0.408	0.251	−0.19	0.567	0.619	0.25	0.45064	0.468	−0.11	−0.047	−0.23	−0.19	0.25	0.354	0.25	
	Complete	0.167	0.931 *	0.891 *	0.655	0.636	0.579	0.7 *	0.858 *	0.799 *	0.8870 *	0.923 *	0.891 *	0.39934	0.514	0.579	0.799 *	0.597	0.799 *	
	Average	0.464	0.941 *	0.858 *	0.888 *	0.852 *	0.418	0.878 *	0.929 *	0.758 *	0.9219 *	0.689	0.858 *	0.61531	0.326	0.418	0.758 *	0.714 *	0.758 *	
	Centroid	0.988 *	0.912 *	0.86 *	0.536	0.801 *	0.492	0.647	0.923 *	0.763 *	0.8988 *	0.656	0.86	0.62657	0.408	0.492	0.763 *	0.733 *	0.763 *	
	Weighted Ward	0.686	0.942 *	0.56	0.194	0.529	0.121	0.527	0.788 *	0.677	0.8746 *	0.893 *	0.56	0.7766 *	0.414	0.121	0.677	0.69	0.677	
		0.167	0.816 *	0.891 *	0.909 *	0.747 *	0.673 *	0.912 *	0.894 *	0.775 *	0.9213 *	0.672	0.891 *	0.8195 *	0.055	0.673	0.775 *	0.479	0.775 *	
Euclidean	Single	0.35	0.788 *	−0.37	0.933	0.626	0.013	0.93 *	0.77 *	0.341	0.03316	−0.08	−0.37	0.28802	0.127	0.013	0.341	0.433	0.341	
	Complete	−0.73	0.694 *	0.767 *	0.893	0.641 *	0.459	0.87 *	0.737 *	0.641	0.07976	0.13	0.767	0.54407	0.576	0.459	0.641	0.594	0.641	
	Average	0.44	0.721 *	0.747 *	0.945 *	0.646 *	0.402	0.926 *	0.758 *	0.663	0.26314	0.557	0.747	0.39126	0.381	0.402	0.663	0.492	0.663	
	Centroid	−0.03	0.933 *	0.767 *	0.952 *	0.608 *	0.363	0.901 *	0.767 *	0.661	0.4001	0.557	0.767	0.26312	0.371	0.363	0.661	0.478	0.661	
	Weighted Ward	0.395	−0.06	0.767 *	0.395	0.478	0.346	0.459	0.206	0.537	0.30413	−0.44	0.767	0.30992	0.127	0.346	0.537	0.358	0.537	
		0.503	0.906 *	0.767 *	0.929 *	0.493	0.45	0.882 *	0.688	0.63	0.53289	0.738	0.767	0.4283	0.38	0.45	0.63	0.571	0.63	
Pearson correlation	Single	0.803 *	0.921 *	−0.51	−0.01	0.918 *	0.351	0.164	0.763	0.358	0.9455 *	0.482	−0.51	0.8374 *	0.68	0.351	0.358	0.597	0.358	
	Complete	−0.69	0.91 *	0.839 *	0.198	0.74	0.629	0.194	0.636	0.549	0.6664	−0.2	0.839	0.16877	0.186	0.629	0.534	0.239	0.549	
	Average	0.703	0.91 *	−0.36	0.117	0.932	0.546	0.363	0.811 *	0.534	0.5964 *	−0.2	−0.36	0.71877	0.556	0.546	0.534	0.677	0.534	
	Centroid	0.177	0.91 *	0.839 *	0.064	0.928 *	0.886 *	0.435	0.82 *	0.822 *	0.8546 *	0.878 *	0.839	0.66499	0.737	0.886 *	0.822 *	0.709	0.822	
	Weighted Ward	0.023	0.91 *	0.839 *	0.03	0.936 *	0.717 *	0.179	0.789 *	0.602	0.79478	0.749	0.839	0.47338	0.164	0.717	0.602	0.319	0.602	
		−0.18	−0.39	−0.55	0.37	0.673	0.289	0.502	0.488	0.329	0.3717	0.749	−0.55	0.58837	0.317	0.289	0.329	0.217	0.329	

* Correlation is relatively large.

4.2. Spatial Patterns of UFZs

According to the results of the hierarchical weighted agglomerative clustering (Figure 6), we found that the hierarchical weighted agglomerative clustering model identified the clusters in an unambiguous way. POI data can be used to identify UFZs to some extent, and the POI data can represent the intensity of human activity. Furthermore, by combining POI and land use type, UFZs can be identified more precisely. Land use type, for example, can be used to identify cultural tourism zones and natural landscape districts. The accuracy and fineness of clustering results were both affected by the number of clusters and segment patches, as shown in Figure 6. Furthermore, the finer the segmentation of the study area, the better the clustering results. To express the spatial autocorrelation of clusters, we used Moran's I index to measure the spatial distribution pattern of the two clustering results. The Moran's I index analysis revealed that the clustering results based on POI and land use type data had a substantial and positive autocorrelation at the 0.05 significance level. The spatial distribution of the clustering results matched that of the actual UFZs (Figure 7). Clustering the segmented sub-regions requires a weighted raw data technique, according to the results. Combining the adjusted cosine distance metric and average clustering linkage strategies can be a suitable method if there is no prior knowledge.

UFZs were identified based on the composition of POI class and land use type data among clusters. Because the region is tiny, some clusters were merged into other clusters. There are seven types of functional zones in downtown Beijing, as shown in Figure 7, including four types of single functional areas and three types of mixed functional zones. The education zone, the recreation green zone, the residence zone, and the social and community zone are single functional zones with areas of 87.4 km², 145.4 km², 153.7 km², and 42.3 km², respectively. With areas of 47.5 km², 73.1 km², and 117.2 km², respectively, the mixed functional areas contain a combination of residence and recreation zones, commercial and industrial zones, and commercial residence zones. The residential zone occupied the most space of all, and it was widely spread out across the study area with significant disparities. On the perimeter, the proportion of the residential zone was higher than in the center. Recreation green zones area denser in the north, but the residential area is relatively far away. Recreation green zones are more dispersed in the south, and they all surround the residential area. As a result, the recreation equity in the south is better. Education zones are concentrated in the Haidian District, i.e., northwest of the study area. Commercial zones are always found in conjunction with other functional zones, such as residence zones, recreation zones, and industrial zones. It also demonstrated that Beijing, as a metropolis, has a relatively effective functional zone plan.

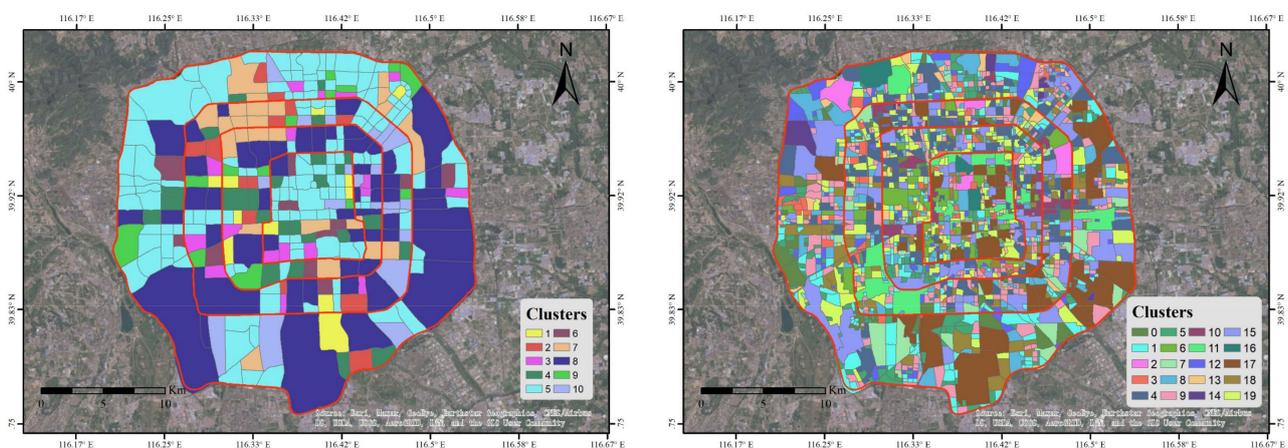


Figure 6. The clustering results when clusters = 10 (left) and clusters = 20 (right) at different segment fineness.

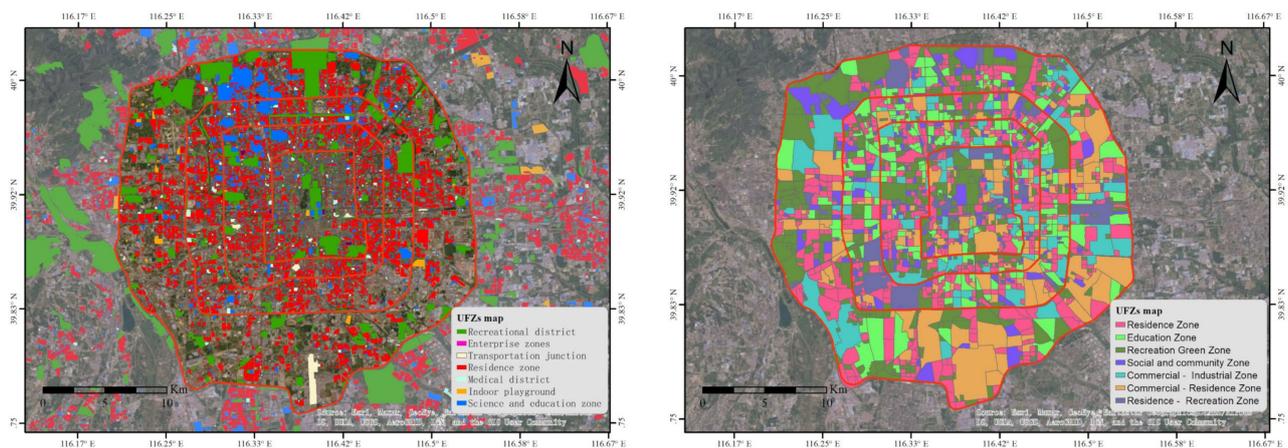


Figure 7. The actual UFZs map from web map (left) and UFZs map from clustering (right).

5. Discussion

5.1. Methodological Advantages and Limitations

The hierarchical weighted clustering model is a popular unsupervised learning technique for discovering the natural groupings of a set of observations, which we used to identify the UFZs. In this study, we proposed a hierarchical weighted clustering model that uses the weighted POI, land use, and socio-economic data to cluster segmented sub-regions divided by road networks. For identifying the urban functional zones, the weights of POI categories scheme, the POI level, distance metrics, and clustering merge strategies were integrated into the clustering model. This study could expand the traditional understanding of clustering based on the individual densities of POIs.

Previous approaches are required to either reduce the raw data into new categories [4,51], which results in the loss of feature information, or simply classify the regions using the raw POI densities [4]. The most significant benefits of our study are that it provides a general paradigm for identifying UFZs and helps to quickly analyze the impact of different characteristic vectors on classification results. Furthermore, we can identify the segmented zones without having any prior knowledge of the label data. Additionally, unlike the K-means algorithm, which has inconsistencies in the results, the hierarchical weighted clustering model could obtain consistent clustering results.

The other advantage is that the entropy weight method was integrated into the evaluation system, making it possible to automatically calculate the weights of hierarchical POI categories. In order to identify UFZs, previous studies usually fail to consider the effects of POI classification level and the weight of POI categories [56]. It has the potential to increase efficiency, unlike the Delphi consensus technique method which requires too much time and money to obtain valuable response through questionnaires. Furthermore, rather than relying on a particular region, it is important to obtain objective and convenient scores in each study area.

There are some limitations to this framework. For example, it is an unsupervised framework, and uncertainty analysis can be problematic due to the lack of prior knowledge in this method. Because the identification of UFZs is based on the feature vectors generated by POIs, some inconsistencies may exist when compared to actual urban functional property. Although this approach has simplified the data processing procedure to consider the weighted and hierarchical relations of POIs, it has yet to establish a uniform mechanism for evaluating the performance of UFZ classification, and all of the processes in this study may need to be repeated in other areas of research.

5.2. Application for Sustainable Urban Planning

The hierarchical weighted clustering model, as opposed to the previous method based on POI density, is clearly more conducive to the analysis and less prone to misinterpretation regarding the weighted and hierarchical relations of POIs. The hierarchical weighted

clustering model is a social-based, planning-oriented, and data-driven classification system linked with the urban function, and it may also be used to connect human activity intensity and UFZ identification. UFZs could identify the heterogeneity of the urban internal thermal environment and quantify the basic units of the effect of anthropogenic heat, as reiterated in a published article on the effects of UHI [15]. The usage of UFZs can provide more precise information than the use of land use and cover data, synonymous with the basic planning unit based on a city's UFZs' pattern. Therefore, the HAC model and the results of UFZs can provide a consistent mapping to urban planning and energy saving inside a city, allowing the UFZs to be applied to city management practices. In general, it is difficult to quantify the impact of human activities on urban heat island effects in an ecological environment because we cannot scientifically partition the intensity of human activities.

This method also has practical significance, and our methodology can advance the understanding of local contexts. For example, the results of the functional zones can be used for identifying the factors of traffic congestion caused by urban planning, analyze the relationship between rainfall water capacity and wettability of small-leaved lime and poplar in different city zones [57], plan a fresh food distribution center based on functional zones for fresh product logistics [58], and provide a means of calibration and reference for urban planning by monitoring the temporal and spatial variability of UFZs [6,59]. Overall, the hierarchical weighted clustering model provides new insights into the methodology of UFZ identification and quantitative assessment of the weight of POI categories, as well as wider application of the impacts of human activities or UFZs on the natural ecological landscape.

6. Conclusions

This study proposed an identification model of UFZs, annotated the social property using POIs and land use data, and provided some potential solutions for the sustainable development of a city on urban functional zones pattern. We found availability and feasibility of hierarchical weighted clustering model. The combination of the adjusted cosine metric and the average criterion revealed the optimal distance metric and linkage strategy, respectively, which has the best performance and quality of clustering results within the Fifth Ring Road, Beijing, China. Compared with the remote sensing images, which primarily depict the physical properties, the results of the clustering model based on POIs data can be viewed as a complementary social sensing view of urban planning and human activities. Despite the fact that semantically meaningful UFZs were identified, the hierarchical weighted clustering model is an unsupervised approach with limits in identifying the actual urban functions. In addition, more research is needed to recognize the social functions accurately while taking into account building height and building density in the study area. This study also provides a valuable method for correlating the natural characteristics and social activities in a densely populated region.

Author Contributions: Conceptualization, R.S.; data curation, Y.J. and R.S.; methodology, Y.J. and R.S.; supervision, R.S.; visualization, Y.J.; writing—original draft, Y.J.; writing—review and editing, Y.J., L.C. and R.S. All authors have read and agreed to the published version of the manuscript.

Funding: The work was financed by the National Natural Science Foundation of China (Grant no. 41922007).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and code presented in this study are available on request from the corresponding author.

Acknowledgments: We thank the Amap™ for providing points of interest data. The land use data used in this study are provided by Gong, P., Tsinghua University. (<http://data.ess.tsinghua.edu.cn/> accessed 1 January 2020).

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

References

1. Lu, Y.; Nakicenovic, N.; Visbeck, M.; Stevance, A.-S. Policy: Five priorities for the UN Sustainable Development Goals. *Nature* **2015**, *520*, 432–433. [[CrossRef](#)] [[PubMed](#)]
2. Zhang, N.; Zhang, J.; Chen, W.; Su, J. Block-based variations in the impact of characteristics of urban functional zones on the urban heat island effect: A case study of Beijing. *Sustain. Cities Soc.* **2021**, *76*, 103529. [[CrossRef](#)]
3. Zhang, X.; Du, S.; Wang, Q. Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data. *ISPRS J. Photogramm. Remote Sens.* **2017**, *132*, 170–184. [[CrossRef](#)]
4. Hu, T.; Yang, J.; Li, X.; Gong, P. Mapping Urban Land Use by Using Landsat Images and Open Social Data. *Remote Sens.* **2016**, *8*, 151. [[CrossRef](#)]
5. Yuan, N.J.; Zheng, Y.; Xie, X. *Discovering Functional Zones in a City Using Human Movements and Points of Interest*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2017; pp. 33–62. [[CrossRef](#)]
6. Zhang, X.; Du, S. A Linear Dirichlet Mixture Model for decomposing scenes: Application to analyzing urban functional zonings. *Remote Sens. Environ.* **2015**, *169*, 37–49. [[CrossRef](#)]
7. Pouyat, R.V.; Trammell, T.L.E. Chapter 10—Climate change and urban forest soils. In *Developments in Soil Science*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 189–211. [[CrossRef](#)]
8. United Nations Department of Economic and Social Affairs. *World Population Prospects: The 2017 Revision, Key Findings and Advance Tables*; United Nations Department of Economic and Social Affairs: New York, NY, USA, 2017.
9. Li, C.; Liu, M.; Hu, Y.; Shi, T.; Qu, X.; Walter, M.T. Effects of urbanization on direct runoff characteristics in urban functional zones. *Sci. Total Environ.* **2018**, *643*, 301–311. [[CrossRef](#)]
10. Güneralp, B.; Reba, M.; Hales, B.U.; Wentz, E.A.; Seto, K.C. Trends in urban land expansion, density, and land transitions from 1970 to 2010: A global synthesis. *Environ. Res. Lett.* **2020**, *15*, 044015. [[CrossRef](#)]
11. Antrop, M. Landscape change and the urbanization process in Europe. *Landsc. Urban Plan.* **2004**, *67*, 9–26. [[CrossRef](#)]
12. Gomes, S.L.; Hermans, L.M. Institutional function and urbanization in Bangladesh: How peri-urban communities respond to changing environments. *Land Use Policy* **2018**, *79*, 932–941. [[CrossRef](#)]
13. Peng, J.; Tian, L.; Liu, Y.; Zhao, M.; Hu, Y.; Wu, J. Ecosystem services response to urbanization in metropolitan areas: Thresholds identification. *Sci. Total Environ.* **2017**, *607–608*, 706–714. [[CrossRef](#)]
14. Du, S.; Du, S.; Liu, B.; Zhang, X. Context-Enabled Extraction of Large-Scale Urban Functional Zones from Very-High-Resolution Images: A Multiscale Segmentation Approach. *Remote Sens.* **2019**, *11*, 1902. [[CrossRef](#)]
15. Yu, Z.; Jing, Y.; Yang, G.; Sun, R. A New Urban Functional Zone-Based Climate Zoning System for Urban Temperature Study. *Remote Sens.* **2021**, *13*, 251. [[CrossRef](#)]
16. Huang, X.; Wang, Y. Investigating the effects of 3D urban morphology on the surface urban heat island effect in urban functional zones by using high-resolution remote sensing data: A case study of Wuhan, Central China. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 119–131. [[CrossRef](#)]
17. Sun, R.; Lü, Y.; Chen, L.; Yang, L.; Chen, A. Assessing the stability of annual temperatures for different urban functional zones. *Build. Environ.* **2013**, *65*, 90–98. [[CrossRef](#)]
18. Peng, J.; Xie, P.; Liu, Y.; Ma, J. Urban thermal environment dynamics and associated landscape pattern factors: A case study in the Beijing metropolitan region. *Remote Sens. Environ.* **2016**, *173*, 145–155. [[CrossRef](#)]
19. Yao, L.; Wei, W.; Yu, Y.; Xiao, J.; Chen, L. Rainfall-runoff risk characteristics of urban function zones in Beijing using the SCS-CN model. *J. Geogr. Sci.* **2018**, *28*, 656–668. [[CrossRef](#)]
20. Ge, M.; Fang, S.; Gong, Y.; Tao, P.; Yang, G.; Gong, W. Understanding the Correlation between Landscape Pattern and Vertical Urban Volume by Time-Series Remote Sensing Data: A Case Study of Melbourne. *ISPRS Int. J. Geo Inf.* **2021**, *10*, 14. [[CrossRef](#)]
21. Su, M.; Zheng, Y.; Hao, Y.; Chen, Q.; Chen, S.; Chen, Z.; Xie, H. The influence of landscape pattern on the risk of urban water-logging and flood disaster. *Ecol. Indic.* **2018**, *92*, 133–140. [[CrossRef](#)]
22. Hou, L.; Wu, F.; Xie, X. The spatial characteristics and relationships between landscape pattern and ecosystem service value along an urban-rural gradient in Xi'an city, China. *Ecol. Indic.* **2020**, *108*, 105720. [[CrossRef](#)]
23. Gao, J.; Yu, Z.; Wang, L.; Vejre, H. Suitability of regional development based on ecosystem service benefits and losses: A case study of the Yangtze River Delta urban agglomeration, China. *Ecol. Indic.* **2019**, *107*, 105579. [[CrossRef](#)]
24. Zhang, H.; Jing, X.-M.; Chen, J.-Y.; Li, J.-J.; Schwegler, B. Characterizing Urban Fabric Properties and Their Thermal Effect Using QuickBird Image and Landsat 8 Thermal Infrared (TIR) Data: The Case of Downtown Shanghai, China. *Remote Sens.* **2016**, *8*, 541. [[CrossRef](#)]
25. Guyot, M.; Araldi, A.; Fusco, G.; Thomas, I. The urban form of Brussels from the street perspective: The role of vegetation in the definition of the urban fabric. *Landsc. Urban Plan.* **2021**, *205*, 103947. [[CrossRef](#)]
26. Pickett, S.T.A.; Cadenasso, M.L. Linking ecological and built components of urban mosaics: An open cycle of ecological design. *J. Ecol.* **2007**, *96*, 8–12. [[CrossRef](#)]
27. Qian, Y.; Zhou, W.; Pickett, S.T.A.; Yu, W.; Xiong, D.; Wang, W.; Jing, C. Integrating structure and function: Mapping the hierarchical spatial heterogeneity of urban landscapes. *Ecol. Process.* **2020**, *9*, 59. [[CrossRef](#)]
28. Herrick, J.E.; Schuman, G.E.; Rango, A. Monitoring ecological processes for restoration projects. *J. Nat. Conserv.* **2006**, *14*, 161–171. [[CrossRef](#)]

29. Shackelford, A.; Davis, C. A combined fuzzy pixel-based and object-based approach for classification of high-resolution multi-spectral data over urban areas. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 2354–2364. [[CrossRef](#)]
30. Cleve, C.; Kelly, M.; Kearns, F.R.; Moritz, M. Classification of the wildland–urban interface: A comparison of pixel- and object-based classifications using high-resolution aerial photography. *Comput. Environ. Urban Syst.* **2008**, *32*, 317–326. [[CrossRef](#)]
31. Ye, S.; Pontius, R.G., Jr.; Rakshit, R. A review of accuracy assessment for object-based image analysis: From per-pixel to per-polygon approaches. *ISPRS J. Photogramm. Remote Sens.* **2018**, *141*, 137–147. [[CrossRef](#)]
32. Li, T.; Cao, J.; Xu, M.; Wu, Q.; Yao, L. The influence of urban spatial pattern on land surface temperature for different functional zones. *Landsc. Ecol. Eng.* **2020**, *16*, 249–262. [[CrossRef](#)]
33. Lan, T.; Shao, G.; Xu, Z.; Tang, L.; Sun, L. Measuring urban compactness based on functional characterization and human activity intensity by integrating multiple geospatial data sources. *Ecol. Indic.* **2021**, *121*, 107177. [[CrossRef](#)]
34. Xu, N.; Luo, J.; Wu, T.; Dong, W.; Liu, W.; Zhou, N. Identification and Portrait of Urban Functional Zones Based on Multisource Heterogeneous Data and Ensemble Learning. *Remote Sens.* **2021**, *13*, 373. [[CrossRef](#)]
35. Yuan, N.J.; Zheng, Y.; Xie, X.; Wang, Y.; Zheng, K.; Xiong, H. Discovering Urban Functional Zones Using Latent Activity Trajectories. *IEEE Trans. Knowl. Data Eng.* **2015**, *27*, 712–725. [[CrossRef](#)]
36. Song, J.; Tong, X.; Wang, L.; Zhao, C.; Prishchepov, A. Monitoring finer-scale population density in urban functional zones: A remote sensing data fusion approach. *Landsc. Urban Plan.* **2019**, *190*, 103580. [[CrossRef](#)]
37. Tu, W.; Hu, Z.; Li, L.; Cao, J.; Jiang, J.; Li, Q.; Li, Q. Portraying Urban Functional Zones by Coupling Remote Sensing Imagery and Human Sensing Data. *Remote Sens.* **2018**, *10*, 141. [[CrossRef](#)]
38. Feng, Y.; Du, S.; Myint, S.W.; Shu, M. Do Urban Functional Zones Affect Land Surface Temperature Differently? A Case Study of Beijing, China. *Remote Sens.* **2019**, *11*, 1802. [[CrossRef](#)]
39. Zimmerbauer, K.; Paasi, A. Hard work with soft spaces (and vice versa): Problematizing the transforming planning spaces. *Eur. Plan. Stud.* **2020**, *28*, 771–789. [[CrossRef](#)]
40. Yuan, J.; Zheng, Y.; Xie, X. Discovering Regions of Different Functions in a City Using Human Mobility and POIs. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; pp. 186–194.
41. Yao, Y.; Li, X.; Liu, X.; Liu, P.; Liang, Z.; Zhang, J.; Mai, K. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 825–848. [[CrossRef](#)]
42. Qian, Z.; Liu, X.; Tao, F.; Zhou, T. Identification of Urban Functional Areas by Coupling Satellite Images and Taxi GPS Trajectories. *Remote Sens.* **2020**, *12*, 2449. [[CrossRef](#)]
43. Song, J.; Lin, T.; Li, X.; Prishchepov, A.V. Mapping Urban Functional Zones by Integrating Very High Spatial Resolution Remote Sensing Imagery and Points of Interest: A Case Study of Xiamen, China. *Remote Sens.* **2018**, *10*, 1737. [[CrossRef](#)]
44. Iranmanesh, A.; Atun, R.A. Reading the urban socio-spatial network through space syntax and geo-tagged Twitter data. *J. Urban Des.* **2020**, *25*, 738–757. [[CrossRef](#)]
45. Alhawarat, M.; Hegazi, M. Revisiting K-Means and Topic Modeling, a Comparison Study to Cluster Arabic Documents. *IEEE Access* **2018**, *6*, 42740–42749. [[CrossRef](#)]
46. Hofmann, T. Probabilistic Latent Semantic Indexing. *ACM SIGIR Forum* **2017**, *51*, 211–218. [[CrossRef](#)]
47. Kiselevich, S.; Mansmann, F.; Nanni, M.; Rinzivillo, S. *Data Mining and Knowledge Discovery Handbook*; Springer: Boston, MA, USA, 2010.
48. Lu, Y.; Wan, Y. PHA: A fast potential-based hierarchical agglomerative clustering method. *Pattern Recognit.* **2013**, *46*, 1227–1239. [[CrossRef](#)]
49. Gong, P.; Chen, B.; Li, X.; Liu, H.; Wang, J.; Bai, Y.; Chen, J.; Chen, X.; Fang, L.; Feng, S.; et al. Mapping essential urban land use categories in China (EULUC-China): Preliminary results for 2018. *Sci. Bull.* **2019**, *65*, 182–187. [[CrossRef](#)]
50. Delgado, A.; Romero, I. Environmental conflict analysis using an integrated grey clustering and entropy-weight method: A case study of a mining project in Peru. *Environ. Model. Softw.* **2016**, *77*, 108–121. [[CrossRef](#)]
51. Wang, Y.; Gu, Y.; Dou, M.; Qiao, M. Using Spatial Semantics and Interactions to Identify Urban Functional Regions. *ISPRS Int. J. Geo Inf.* **2018**, *7*, 130. [[CrossRef](#)]
52. Lessig, V.P. Comparing cluster analyses with cophenetic correlation. *J. Mark. Res.* **1972**, *9*, 82–84. [[CrossRef](#)]
53. Saraçlı, S.; Doğan, N.; Doğan, I. Comparison of hierarchical cluster analysis methods by cophenetic correlation. *J. Inequalities Appl.* **2013**, *2013*, 203. [[CrossRef](#)]
54. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
55. Contreras-Vidal, J.L.; Kerick, S.E. Independent component analysis of dynamic brain responses during visuomotor adaptation. *NeuroImage* **2004**, *21*, 936–945. [[CrossRef](#)]
56. Liu, X.; Long, Y. Automated identification and characterization of parcels with OpenStreetMap and points of interest. *Environ. Plan. B Plan. Des.* **2015**, *43*, 341–360. [[CrossRef](#)]
57. Klamerus-Iwan, A.; Błońska, E.; Lasota, J.; Waligórski, P.; Kalandyk, A. Seasonal variability of leaf water capacity and wettability under the influence of pollution in different city zones. *Atmos. Pollut. Res.* **2018**, *9*, 455–463. [[CrossRef](#)]

-
58. Wu, C.; Hu, T.J.; Wang, X.F.; Zheng, C. Study on the Functional Zones Layout of Fresh Food Distribution Center Based on the SLP Method. *Adv. Mater. Res.* **2013**, *694*, 3614–3617. [[CrossRef](#)]
 59. Gao, S.; Janowicz, K.; Couclelis, H. Extracting urban functional regions from points of interest and human activities on location-based social networks. *Trans. GIS* **2017**, *21*, 446–467. [[CrossRef](#)]