Article

Mapping Urban Land Use by Using Landsat Images and Open Social Data

Tengyun Hu [1], Jun Yang [1,2,*], Xuecao Li [1] and Peng Gong [1,2]

[1] Ministry of Education Key Laboratory for Earth System Modeling, Center for Earth System Science, Tsinghua University, Beijing 100084, China; hutengyun88@163.com (T.H.); li-xc12@mails.tsinghua.edu.cn (X.L.); penggong@tsinghua.edu.cn (P.G.)
[2] Joint Center for Global Change Studies, Beijing 100875, China
* Correspondence: larix@tsinghua.edu.cn

Abstract: High-resolution urban land use maps have important applications in urban planning and management, but the availability of these maps is low in countries such as China. To address this issue, we have developed a protocol to identify urban land use functions over large areas using satellite images and open social data. We first derived parcels from road networks contained in Open Street Map (OSM) and used the parcels as the basic mapping unit. We then used 10 features derived from Points of Interest (POI) data and two indices obtained from Landsat 8 Operational Land Imager (OLI) images to classify parcels into eight Level I classes and sixteen Level II classes of land use. Similarity measures and threshold methods were used to identify land use types in the classification process. This protocol was tested in Beijing, China. The results showed that the generated land use map had an overall accuracy of 81.04% and 69.89% for Level I and Level II classes, respectively. The map revealed significantly more details of the spatial pattern of land uses in Beijing than the land use map released by the government.

Keywords: urban land parcel; remote sensing; social data; land use

---

1. Introduction

Urbanization in China is taking place at a fast rate [1]. The proportion of urban population increased from 20% in 1982 to more than 50% in 2014 [2]. Large-scale urbanization has had a dramatic impact on the environment and the wellbeing of seven hundred million urban residents. Studies that assess this process and its impacts are important for taking remedial actions and designing better urbanization strategies for the future. To achieve these goals, detailed urban land cover/use maps are required.

Currently, land cover information with resolutions ranging from low to high is the primary data source used in studies such as urban growth simulation [3–5], evaluation of urban public health [6], and assessment of urban ecosystem services [7–9]. However, to study issues such as housing provision [10,11], urban transportation, job accessibility and residential relocation [12,13], and land use patterns [14], detailed information on urban land use is needed due to the difference between the two concepts: land use is a cultural concept that describes human activities and their use of land, whereas land cover is a physical description of land surface [15]. Land cover can be used to infer land use, but the two concepts are not entirely interchangeable.

Nevertheless, high-resolution urban land use maps covering large spatial extents are relatively rare because local knowledge and the techniques necessary for developing these types of maps are often not available, particularly for developing regions [16,17]. Moreover, urban land use maps are

normally produced by interpreting aerial photographs, field survey results, and auxiliary materials, such as appraisal records or statistical data [18,19]. The evolving nature of urban development often outruns the on-and-off efforts to update existing land use databases and results in outdated maps. To make the situation worse, high-resolution land use maps are frequently kept out of the reach of the general public. As a result, to obtain land use maps that capture the pace of urban development in a timely and accurate manner at a relatively large spatial scale is a critical challenge in urban studies, both in China and in other countries facing similar situations.

Satellite-based remote sensing holds certain advantages in monitoring the dynamics of urban land use because of the large spatial coverage, high time resolution, and wide availability. Pixel-based image classification methods using spectral [20] and/or textural properties [21,22] are frequently applied to extract urban land use information. Recently, per-field and object-based classification methods have gained popularity in deriving land uses from the satellite images [19,23] because per-field classification methods can better describe the function of urban areas and serve the needs of urban planning [24]. Although significant progress has been achieved, deriving high-resolution urban land use maps from satellite images is still a difficult task. The medium-resolution satellite images (e.g., Landsat images) allow for mapping urban areas at the large spatial scale, but it is still difficult to extract socioeconomic features of urban areas from these images [25]. Land cover information derived from medium-resolution satellite images cannot provide sufficient separation among urban functional zones. Satellite images with high spatial and spectral resolution provide more detailed information on urban structures and thus facilitate the assignment of socioeconomic functions to different zones. Nevertheless, these images are prohibitively expensive in general.

The emergence of open social data creates new opportunities for mapping urban land uses at high-resolution. Open social data containing spatiotemporal patterns of human activities can be used to uncover land uses and intra-city functions [25]. Efforts have been made to use social data sets, such as mobile phone records [26–28], taxi data [29,30], and smart card data [31], to reveal characteristics of urban land uses. The results have shown that social data can accurately capture the spatiotemporal rhythm of human activities, which provides a way for exploring how cities function at fine spatiotemporal resolution [32]. However, the existing studies were often implemented over relatively small areas or specific land use types using data sets that were subjective or proprietary. In addition, the physical attributes of urban functional parcels, which can help determine land use at the parcel level, were seldom included in these studies [25].

There is strong potential to combine the strength of these two data sources, *i.e.*, integrating social knowledge with remotely sensed data, to gain better insights into urban land use patterns. Physical features extracted from satellite data and socioeconomic features retrieved from open social data can be combined to characterize urban land uses more accurately. One type of open social data that is particularly promising for this purpose are Points of Interest (POI) data. POI data are geographical data provided voluntarily by individuals. The data are mainly used for monitoring users' positions in spatial tracking or geo-caching systems. POI data can link geographic locations to particular places, descriptive features, and other place-based information. In other words, POI data can link the informal world of everyday human discourse and the formal world of GIS (Geographic Information System) [33–36]. As far as we know, there are no reports on using POI data and satellite data jointly to produce detailed land use maps. To fill in this gap, we developed a protocol that utilizes medium-resolution satellite images and POI data to map urban land uses. In the rest of the paper, we describe the details of the protocol and its application in Beijing, China. We also discuss the strengths and limitations of the proposed protocol and make suggestions for future work in this field.

2. Study Area and Dataset

2.1. Study Area

Beijing, the capital of China, is located in the North China Plain. The south and east parts of the city are plains, whereas the remaining areas are surrounded by mountains. Beijing has an administrative area of 16,808 km², which includes a mix of urban and rural land uses. The urban and rural land use types are also referred to as built-up (constructed or developed) regions and non-built-up regions in China. The area of the built-up regions is approximately equivalent to the total impervious surface area of the entire region (Figure 1).
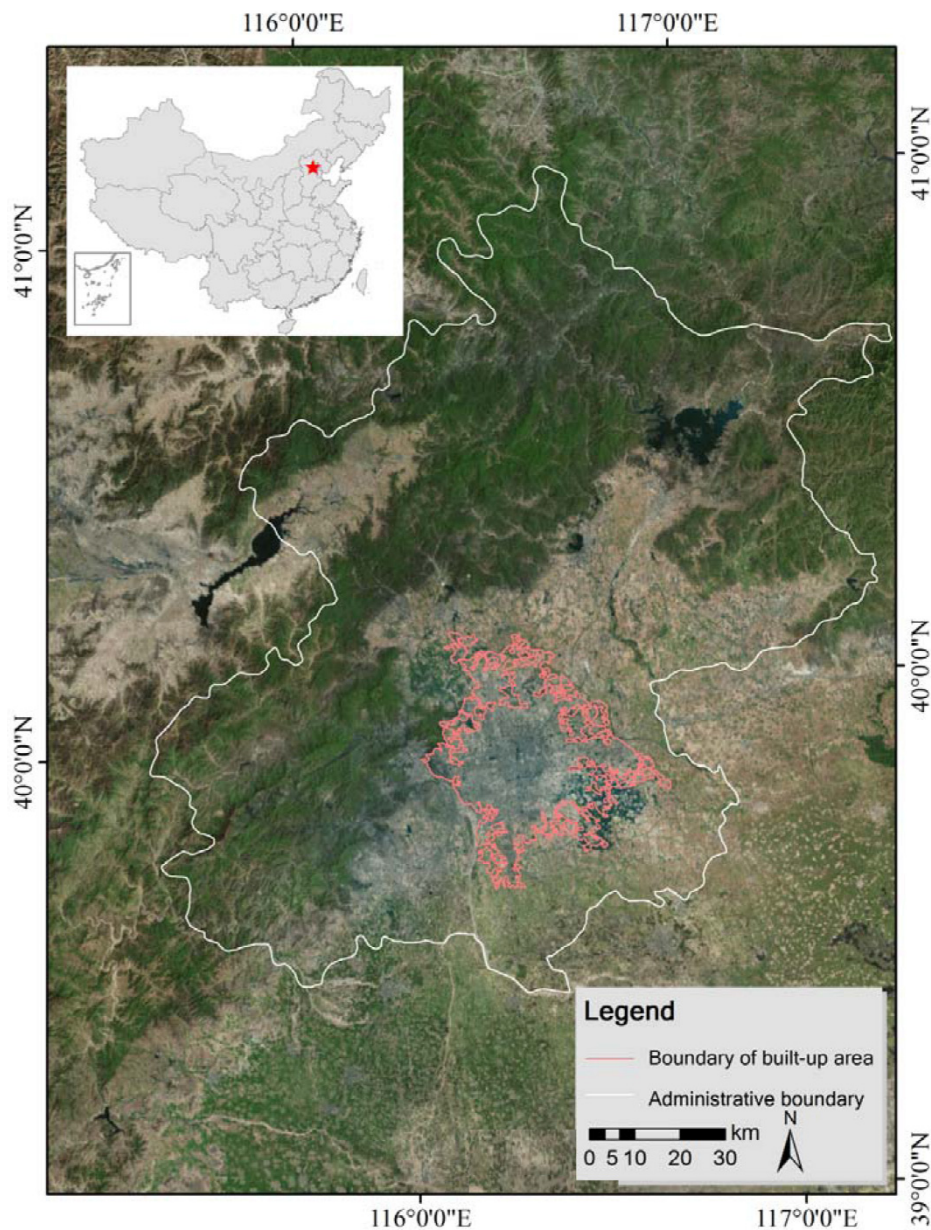
Figure 1. Map of the study area. The white line shows the boundary of the administrative area of Beijing, China. The built-up areas are at the center of the satellite image and are indicated by a red boundary.

2.2. Data Collection

The administrative boundary of Beijing falls over two sets of Landsat images (path/row: 123/32 and 123/33). Totally 14 Landsat 8 Operational Land Imager (OLI) images of 2013 were procured as our primary data source from the U.S. Geological Survey (http://earthexplorer.usgs.gov/). These images were selected because of the low cloud proportions (<10%). Multiple available Landsat images with good quality in 2013 were used to remove the impact of cloud contamination, phenology of vegetation and cropland rotation. In addition, a seasonal dynamic of land cover series, generated by Wang *et al.* [37], was used to help identify urban parcels.

Data on the road networks of Beijing were collected from Open Street Map (OSM) (https://www.openstreetmap.org), a provider of free open geographical data. The data are in vector format and contain different classes of streets organized using street levels and sizes. Street levels, in descending order, correspond to primary highways, primary roads, secondary roads, and small roads (*i.e.,* local, neighborhood and rural streets).

Also more than 30 million POIs in Beijing were gathered from http://www.datatang.com/data/44484. These points were acquired from a popular social network in China (SINA WEIBO, a social network similar to Twitter), through a location-based service. Each point contains the functional and locational properties of a site, which has been automatically recognized by mainstream map suppliers (*i.e.,* Google or Baidu maps) [38]. The initial twenty types of POI were aggregated into 10 general categories, including residential, marketing and recreation, service building, hotel and restaurant, industrial, medical, educational, institutional infrastructure, government and social organization, and transportation land (see Table S1). POIs that did not belong to the aforementioned groups were removed. The quality of the POI data were verified by checking 100 randomly sampled sites manually, and the resulting accuracy level was 97%. Although spurious social data may occur, the overall pattern (or distribution) can be accurately reflected by using a huge amount of points.

3. Method

The overall structure of the protocol is shown in Figure 2. First, the entire study area was segmented into parcels based on road networks following the methods developed by Long and Liu [38] (Figure 2-1). Parcels are basic units used in this classification scheme with the assumption that they are homogeneous in terms of urban functions [39]. The parcels were then separated into built-up areas and non-built-up areas based on classified impervious surface areas [37] and defined our classification system based on these two regions (Figure 2-2). The function of each parcel was inferred using the normalized feature distance (or similarity) to the pre-collected training sample units. The similarity of the built-up parcels was based on 10 socioeconomic features (*i.e.,* residential, marketing and recreation, service building, hotel and restaurant, industrial, medical, educational, institutional infrastructure, government and social organization and transportation land) that were derived from the normalized kernel densities of the different functions of POI data and two physical indices derived from multi-temporal Landsat images (Figure 2-3). The land use types of the non-built-up parcels were determined by the dominant land cover types of the parcels. Finally, the classified built-up and non-built-up regions were merged into a final land use map of the city and assessed its accuracy (Figure 2-4).
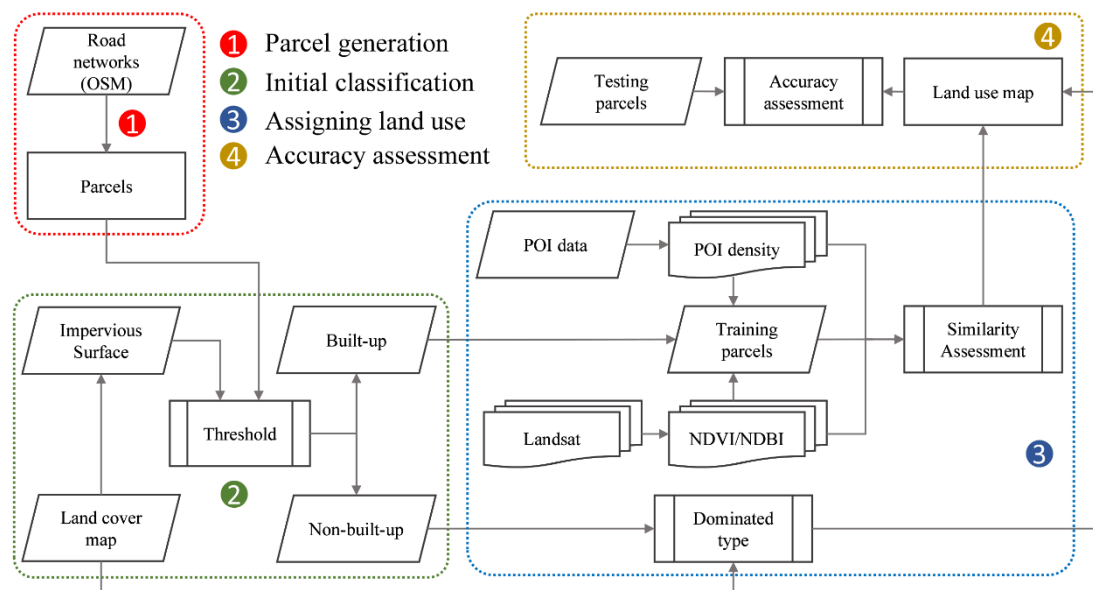
**Figure 2.** The proposed flowchart for mapping detailed land use.

3.1. Parcel Generation

A parcel is the basic unit carrying socioeconomic functions in urban management and urban planning. A parcel is relatively homogeneous in terms of land use function [40]. Long and Liu [38] suggested that parcels are polygons bounded by road networks, which serve as natural segmentation boundaries of the urban area. In this study, we adopted this assumption and defined parcels using OSM road networks data. To remove some unnecessary details (e.g., overpasses), first a trim operation was performed for roads shorter than 500 m and both ends of roads for 100 m were extended to form road segments by connecting originally non-connected lines. As a result, small holes were filled and some incomplete roads were removed. Then, road spaces were generated through a dilation operation around road networks by referring to MoHURD (Ministry of Housing and Urban-Rural Development) [41] and the actual road situation in Beijing. Different road widths were used for different road levels, e.g., 55 m, 50 m, 40 m, and 35 m for the first, second, third and fourth levels, respectively. Finally, land parcels bounded by roads were generated after removing the road spaces (see Figure 3). After parcel generation, 17692 parcels of built-up area and 11072 parcels of non-built-up area were obtained.

3.2. Classification System

MoHURD defined the built-up area as places dominated by artificial buildings and structures and non-built-up area as places mainly occupied by cultivated land, forests, grassland, water and water conservancy facilities [42–44]. First the proportion of impervious area for each parcel was computed based on the classified impervious surface map in the land cover map product [37]. In this study, a threshold of 0.3 were used as suggested by [45,46], to differentiate the built-up and non-built-up areas. If the proportion of impervious surface of a parcel was higher than the threshold, then this parcel was designated as built-up. Otherwise, the parcel was labeled as non-built-up.

For the non-built-up area, the classification system of land cover map developed by Gong *et al.* [47] was adopted, and the description is shown in Table 1. Referring to the Chinese land use classification criteria (GB/T21010-2007), the built-up areas were divided into four Level I classes and nine Level II classes (see Table 1).
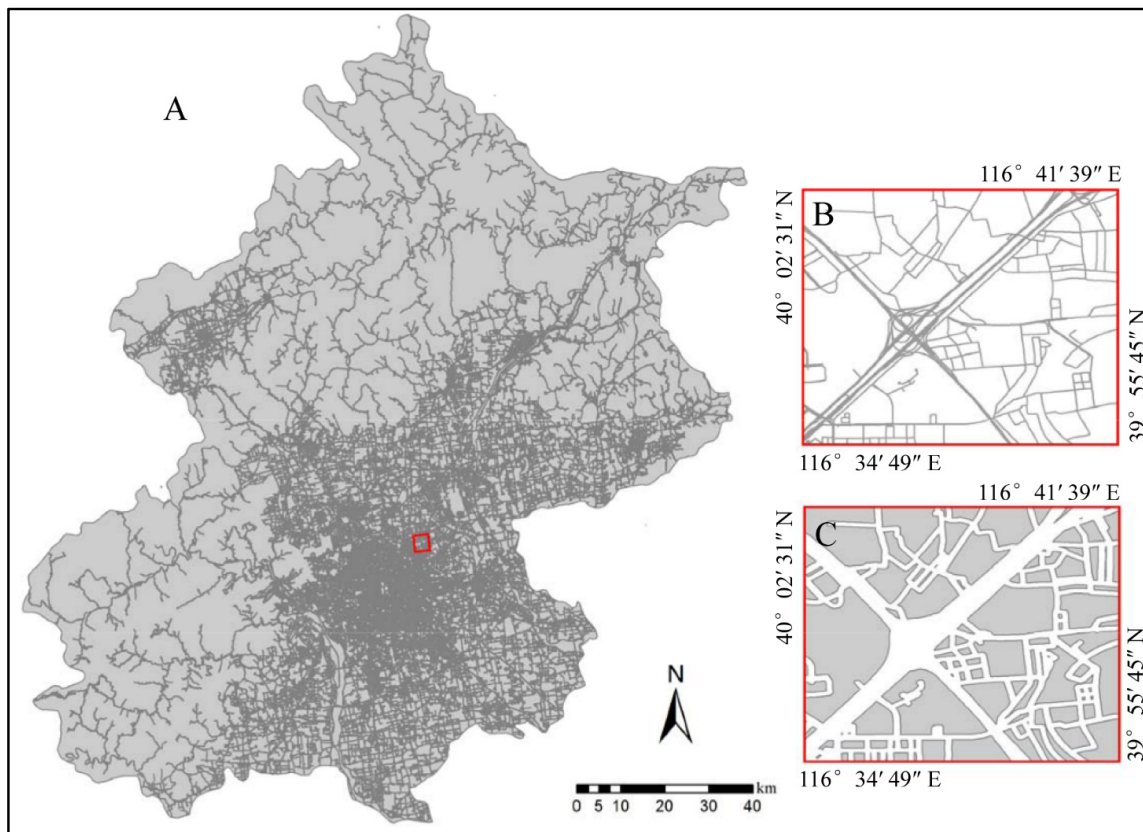
Figure 3. Distribution of urban land use parcels. (**A**) Overall pattern; (**B**) and (**C**) are zoomed-in views (red frame) of the original road networks and the segmented land parcels.

Table 1. Land use classification system.

Land Use Boundary	Level I Classes	Level II Classes	Descriptions
Non-built-up region (refer to Gong *et al.* [47])	Agriculture	Cropland	Farm lands with human activities
		Orchard	Parcels planted with fruit trees or shrubs
	Green Space	Forest	Trees with distinct canopy textures
		Grassland	Grasslands for grazing
		Shrubland	Shrubs with textures finer than trees but coarser than grasslands
	Waterbody	Waterbody	Natural or artificial waterbodies
	Undeveloped	Undeveloped	Vacant land, bare land and land under construction
Built-up region	Residential	Cottage	Small, single-story houses (rural settlements, traditional dwellings *i.e.,* Hutong, Siheyuan)
		Community	Residential areas with a protective gate and detailed architectural standards (e.g., floor area ratio and greening rate)
	Commercial	Retail place	Retail and general merchandise, apparel and accessories, grocery and food sales, eating and drinking, hotel, and recreational facilities (e.g., cinemas and amusement parks)
		Service building	Mixed function (architecture combining commercial with residential functions), financial services (banks), office buildings
	Industrial	Industrial land	Manufacturing, warehousing, equipment sales and service, recycling and scrap, mining facilities
	Institutional	Medical place	Hospitals
		Educational/Research place	Educational facilities
		Administrative offices	Government services, diplomatic buildings, military facilities
		Public service	Railroad facilities, transportation terminals, aviation facilities, meeting and assembly facilities, municipal utilities

3.3. Type Identification

3.3.1. Training Parcels

For land use mapping in built-up areas, training parcels for Level II were selected by referring to high-resolution images on Google Earth and Baidu Map (http://map.baidu.com/) [47]. Selected training samples were the most typical representatives of the Level II classes. For instance, the class of cottage is usually distributed in suburban areas and has higher housing intensity than other classes. The community type features a protective gate and some green vegetation. The class of industrial use contains areas for manufacturing, assembly, and fabrication operations, often over a relatively large scale but with lower intensity [23]. For the class of commercial use, the floor area ratio and building density are higher than industrial and residential uses. A typical characteristic of institutional use is that parcels usually contain a lower housing intensity than that of residential use or commercial use [19]. Institutional use is designated for specific public services, where the size of the site is adequate for its use and is also highly accessible to roads. Based on these basic features, as well as the field survey, 120 training sample units at the parcel level were collected, in which we are confident in terms of their representativeness. Table 2 lists the number of training parcels adopted in the built-up regions. Figure 4 shows some examples of the collected training sample units.

Table 2. Quantity of collected training parcels.

| Level I	Num.	Level II	Num.
Residential	25	Cottage	7
		Community	18
Commercial	39	Consuming	22
		Services	17
Industrial	14	Industrial	14
Institutional	42	Medical	10
		Educational	11
		Administrative	8
		Public	13

3.3.2. Processing POIs

Within the spatial extent of a parcel, there may be a variety of POIs of different types, which can be regarded as having compound functions instead of a single function [40]. In addition, the qualities of POIs vary among different categories, *i.e.,* the number of POIs associated with the commercial type is greater than the other types. This results in an unbalanced distribution of the numbers of points among different POI types. To cope with these issues, we normalized the functional intensity of the different POI types using kernel density estimation. Kernel density analysis was implemented using the quadratic kernel function [48] with a search radius of 500 m. The output is a smooth surface indicating the densities, and regions with relatively higher density values indicate that there are more POI points. This processing can mitigate possible errors caused by unbalanced quantity gaps among different POI types. Figure 5 shows the density images for the different functional POIs (Table S1) adopted in this study. To facilitate comparison among different functional POIs, a normalization procedure was implemented in advance to remove the different quantities of POIs among the different function types (see Equation (1)).

$$V_{norm} = \frac{V - V_{min}}{V_{max} - V_{min}} \tag{1}$$

where V_{norm} is the normalized value of the POI density map, and V_{min} and V_{max} are the minimum and maximum values in the POI density maps, respectively.
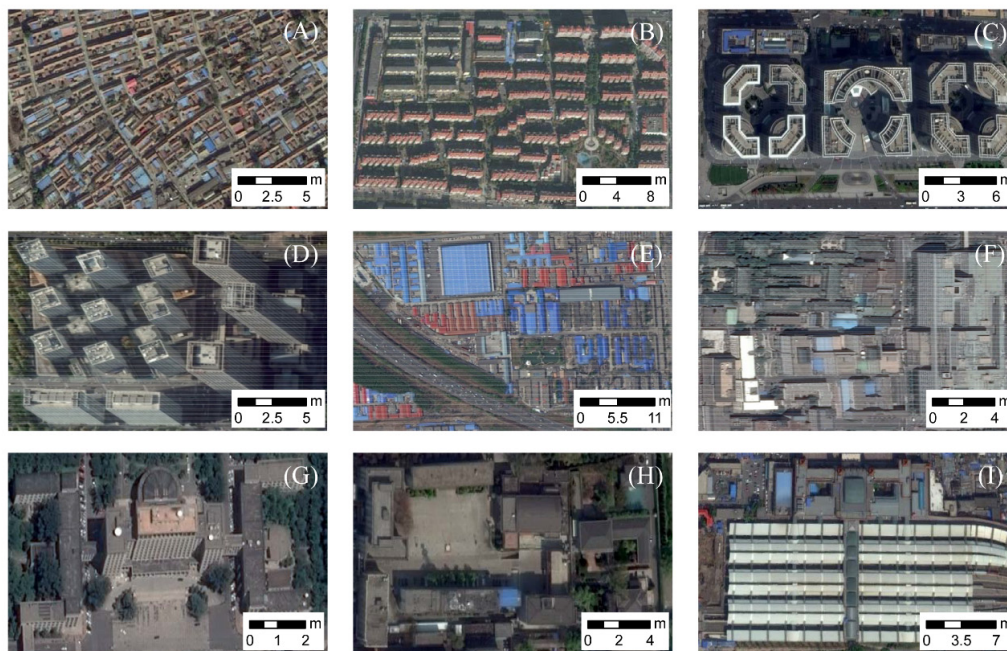
Figure 4. Training samples for nine subclasses of land use in the built-up regions: (**A**) cottage; (**B**) community; (**C**) retail place; (**D**) service building; (**E**) industrial land; (**F**) medical place; (**G**) educational/research place; (**H**) administrative office; (**I**) public service.
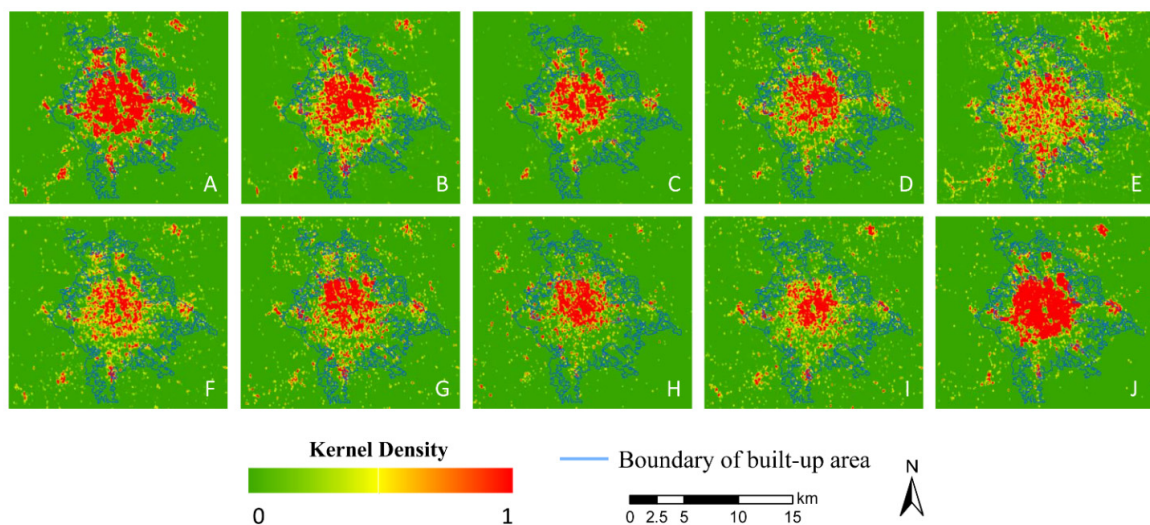
Figure 5. Normalized Kernel density maps of Points of Interest (POI) data: (**A**) residential; (**B**) marketing and recreation; (**C**) service building; (**D**) hotel and restaurant; (**E**) industrial; (**F**) medical; (**G**) educational; (**H**) institutional infrastructure; (**I**) government and social organization; and (**J**) transportation land.

3.3.3. Retrieving Physical Features

To measure the quantity of green vegetation and building in each parcel, the Normalized Difference Vegetation Index (NDVI) and Normalized Difference Built-up Index (NDBI) were derived from Landsat images. NDVI is calculated using Equation (2). Several Landsat OLI images were used with acceptable image quality in 2013 to remove the impact of cloud contamination, phenology of vegetation, and cropland rotation [49]. Multi-temporal NDVI indices were used to generate a composite maximum NDVI (*i.e.*, $NDVI_{max}$) in 2013 (see Equation (3)).

$$NDVI = (NIR - RED) \, / \, (NIR + RED) \tag{2}$$

where *NIR* is band 5 of Landsat 8, and *RED* is band 4 [50].

$$NDVI_{max} = MAX \, [NDVI_1, NDVI_2, ..., NDVI_n] \tag{3}$$

where $NDVI_1, NDVI_2, ..., NDVI_n$ are the multi-temporal NDVI indices in 2013.

　　Similarly, NDBI is calculated using Equation (4), where MIR is OLI band 6. The index was useful for detecting impervious surfaces [51]. As NDBI is a relatively stable index within one year, it was calculated using only one image acquired on 3 October 2013 because of its low cloud coverage.

$$NDBI = (MIR - NIR) \, / \, (MIR + NIR) \tag{4}$$

3.3.4. Determination of Parcel-Based Land Use

　　Using the training parcels that were collected based on the definition of the land use classification system given in Table 1, a normalized feature distance (similarity) in built-up regions was computed on a parcel-by-parcel basis (Equation (5)). The features used for calculating the similarity index include 10 POI density images, one NDVI band and one NDBI band. Two statistical parameters (*i.e.,* the mean value and standard deviation) were previously estimated using the collected training samples. Then, the similarity index of a given parcel was compared with each type [52,53].

$$S_i = \sqrt{\frac{1}{m} \sum_{j=1}^{m} \left(\frac{x_j - \overline{x_i}}{\sigma_i} \right)^2} \tag{5}$$

where $\overline{x_i}$ and σ_i are the mean and standard deviation of the pre-defined land use type *i* acquired from the training parcel; *m* is the total number of land use classes; and x_j is the parcel value (*i.e.,* the mean of all pixels within the parcel) for feature *j* in either normalized POI density images, NDVI or NDBI. The smaller feature distance means higher similarity to a corresponding land use type. The urban land use type of a parcel was determined by calculating similarity S_i of the parcel to training samples, and the pre-defined land use type of training samples which has the minimum value of S_i was assigned to this parcel.

　　In addition, the land cover map was adopted to determine the land use of parcels in non-built-up lands. A land parcel commonly has multiple land cover types. The dominant land cover type was identified and the land use function of this land cover type was assigned to the corresponding parcel. Finally, the classified built-up and non-built-up areas were combined to form the detailed land use map for the entire city.

*3.4. Accuracy Assessment and Uncertainty*

　　To assess the performances of land use classification, a random sampling scheme was adopted to collect a testing sample set over the study area [54]. All testing parcels were surveyed by a field crew with a relatively high level of confidence. The total number of collected testing parcels was 269, among which 180 were located in built-up regions, and 89 were in non-built-up areas. Thereafter, confusion matrixes for Level I and Level II were built.

　　To assess the uncertainty of the obtained urban land use map, the standard deviation of similarities among all land use types for each parcel was used as an approximate indicator. Low standard deviation values suggested that there was a relatively little similarity difference among the different types of land use, which meant that the identified land use was more uncertain. High standard deviation values indicated considerable variance of similarities among different land use types, which would tend to result in a more credible assigned type of land use to a parcel because the minimum similarity was considerably different. This approach qualitatively captured the overall pattern of the uncertainty distribution.

4. Results

4.1. Performance of the Land Use Map in Beijing

The generated land use map of Beijing is shown in Figure 6 (*i.e.,* A and C represent the overall and local views, respectively). This map contains eight Level I classes and 16 Level II classes, and their confusion matrixes are given in Tables 3 and 4 respectively. We obtained an overall accuracy (OA) of 81.04% and a kappa coefficient of 0.78 for the eight Level I classes (Table 3). For the Level II classes (Table 4), the OA and kappa coefficient were 69.89% and 0.68%, respectively. For Level I, the residential land use class had the highest producer's accuracy (85%) in built-up areas, whereas the type of institutions had a relatively lower value of 66%. However, the categories with the highest producer's accuracy were community (74%) and industrial areas (74%) in the Level II classes, and the lowest was administrative departments (33%).
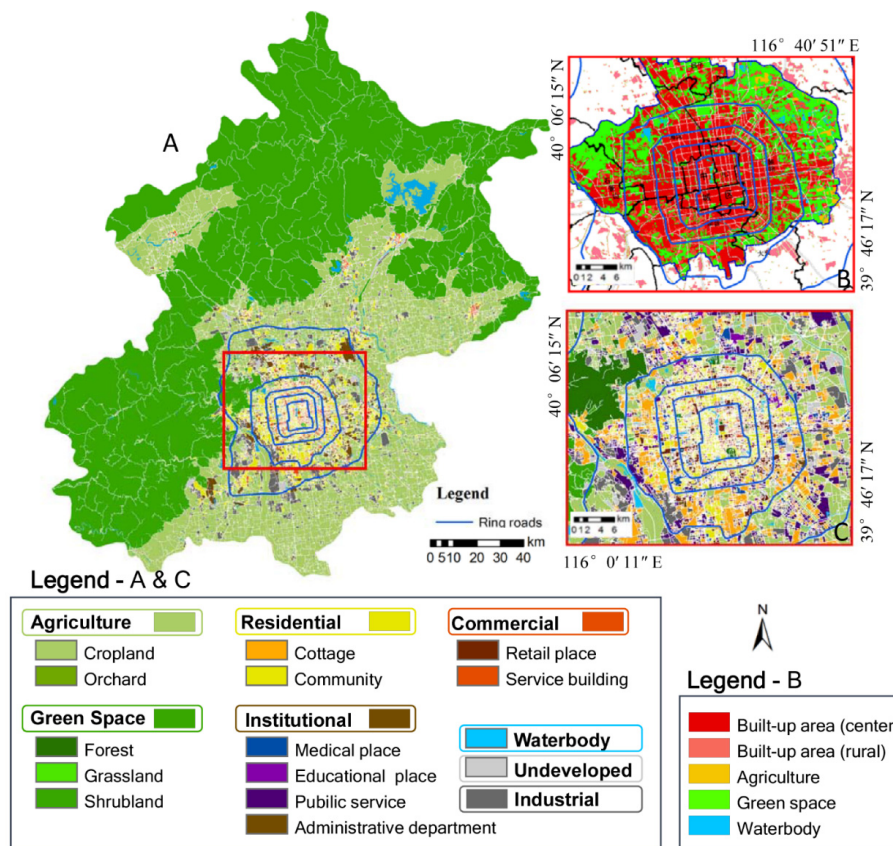
Figure 6. Detailed land use map of the Beijing area in 2013. (**A**) Overview map of the Level I land use; (**B**) zoomed-in view of the official land use map in Beijing; (**C**) detailed map of Level II land use types with the same extent of (**B**).

Table 3. Confusion matrix of the classification of Level I urban land use types.

Level I (Abbr)	A	G	W	U	R	C	Ind	Ins	Total	UA(%)
Agriculture (A)	**25**	0	0	0	1	0	0	0	26	96%
Green Space (G)	0	**29**	1	3	0	0	0	0	33	88%
Water Body (W)	1	0	**18**	0	0	0	0	0	19	95%
Undeveloped (U)	0	0	0	**11**	0	0	0	0	11	100%
Residential (R)	0	0	0	0	**47**	7	3	7	64	73%
Commercial (C)	0	0	0	0	2	**32**	1	6	41	78%
Industrial (Ind)	0	0	0	0	1	0	**14**	9	24	58%
Institutional (Ins)	0	0	0	0	4	4	1	**42**	51	82%
Total	26	29	19	14	55	43	19	64	269	
PA(%)	96%	100%	95%	79%	85%	74%	74%	66%		**81.04%**

Table 4. Confusion matrix of the classification of Level II urban land use types.

Level II (Abbr)	W	Cro	Orc	For	Gra	Shr	U	Cot	Com	Ret	Ser	Ind	Med	Edu	Adm	Pub	Total	UA(%)
Water (W)	**18**	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	95%
Cropland (Cro)	0	**15**	1	0	0	0	0	0	1	0	0	0	0	0	0	0	17	88%
Orchard (Ora)	0	0	**9**	0	0	0	0	0	0	0	0	0	0	0	0	0	9	100%
Forest (For)	0	0	0	**14**	0	3	1	0	0	0	0	0	0	0	0	0	18	78%
Grassland (Gra)	0	0	0	0	**6**	0	0	0	0	0	0	0	0	0	0	0	6	100%
Shrubland (Shr)	1	0	0	1	0	**5**	2	0	0	0	0	0	0	0	0	0	9	56%
Undeveloped (U)	0	0	0	0	0	0	**11**	0	0	0	0	0	0	0	0	0	11	100%
Cottage (Cot)	0	0	0	0	0	0	0	**13**	3	1	2	1	0	0	0	0	20	65%
Community (Com)	0	0	0	0	0	0	0	8	**23**	3	1	2	1	1	3	2	44	52%
Retail (Ret)	0	0	0	0	0	0	0	1	1	**11**	4	0	1	2	0	0	20	55%
Service (Ser)	0	0	0	0	0	0	0	0	0	4	**13**	1	1	1	1	0	21	62%
Industrial (Ind)	0	0	0	0	0	0	0	0	1	0	0	**14**	5	1	1	2	24	58%
Medical (Med)	0	0	0	0	0	0	0	0	1	0	0	0	**8**	0	0	0	9	89%
Educational (Edu)	0	0	0	0	0	0	0	0	2	0	1	1	1	**12**	3	1	21	57%
Administrative (Adm)	0	0	0	0	0	0	0	0	0	0	1	0	1	0	**4**	0	6	67%
Public (Pub)	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	**12**	15	80%
Total	19	16	10	15	6	8	14	23	32	20	23	19	18	17	12	17	269	
PA(%)	95%	94%	90%	93%	100%	63%	79%	57%	72%	55%	57%	74%	44%	71%	33%	71%		**69.89%**

　　　We also compared the only land use map that is available to the general public from the Beijing Municipal Bureau of Land and Resources (http://www.bjgtj.gov.cn/art/2010/1/8/art_2340_90537.html) with our result (Figure 6B,C). The mapping year of the official data is 2005 (scale, 1: 100,000). The land use map generated in this study revealed much more information on urban land use patterns in central Beijing. The official map only shows the distribution of built-up areas and non-built-up areas (including agricultural areas, water bodies and green space).

4.2. The Land Use Pattern of Beijing

　　　Given that the spatial structure of Beijing bears an approximate radial concentric pattern defined by ring roads [55], we calculated the percentages of Level I land use classes of areas contained in the ring roads (Table 5). The built-up areas were mainly distributed inside the 5th ring road, and the largest portion was the class of residential land use, which accounted for more than 40% of the built-up areas. More than 80% of the commercial land use was located inside the 4th ring road. In contrast, the percentage of industrial land use was the highest between the 5th and 6th ring roads. For the non-built-up areas, the class of agriculture land use was mainly distributed outside the 5th ring road and accounted for 67.61% of the total area. Green space was the largest land use class outside of the 6th ring road, with a value of 67.94%.

Table 5. Percentages of different land use types among the ring roads.

Percentage (%)	Inside 2nd	2nd–3rd	3rd–4th	4th–5th	5th–6th	Outside 6th
Agriculture	0.98	0.73	5.69	17.53	40.46	27.15
Green space	7.36	2.33	0.52	1.79	9.65	67.94
Water body	7.34	2.33	1.13	1.64	1.81	1.31
Undeveloped	11.24	8.55	3.38	3.51	3.89	0.35
Residential	41.40	48.98	46.70	43.47	15.52	1.00
Commercial	28.71	32.20	23.75	8.04	1.05	0.09
Industry	0.35	0.58	3.70	6.46	14.28	1.56
Intuitional	2.62	4.29	15.13	17.55	13.34	0.60

　　　According to Figure 6 and Table 5, the urban core area mainly provided commercial and political functions. Groups of commercial and institutional parcels were distributed inside the 4th ring road and helped to form a multi-nuclei development pattern outside the urban core area. Residential areas were distributed the most widely of all land use types as a result of the large-scale development of residential blocks to meet the high housing demands since the early 1980s [49]. The industrial areas were mainly concentrated at the urban fringe, probably due to less stringent environmental standards and the availability of cheap land in this region.

4.3. Uncertainty of the Mapped Result

　　　The standard deviation of similarity based on each parcel was used to assess the confidence level of the identified land use type (Figure 7). The standard deviations of the parcels were relatively high inside the 4th ring road and gradually decreased from the center to the outside. We were more confident with the identified land use types in the center part of the city due to the relatively small areas of parcels and the higher abundance of POIs in this region.
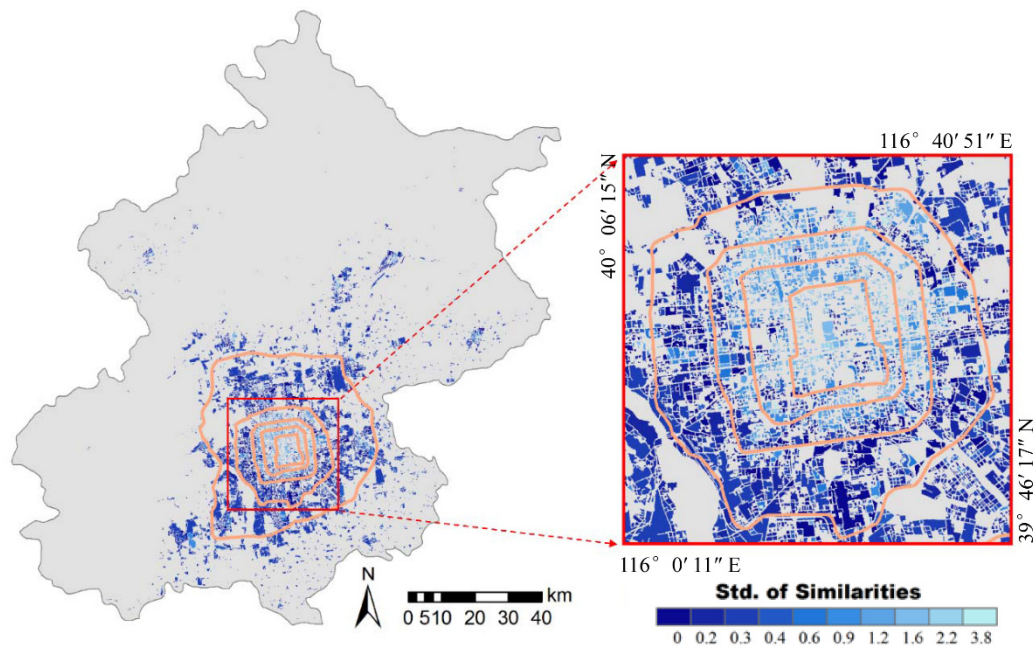
Figure 7. Map of the standard deviation of the similarities of parcels. The 2nd to the 5th ring roads are shown by orange lines.

5. Discussion

In this study, we utilized both medium-resolution satellite remote sensing data and open social data to identify detailed land use classes in urban areas. The global pattern of urban land use distribution is well-reflected. Physical features (*i.e.,* NDVI and NDBI) derived from remote sensing data describe the biophysical elements of urban areas. The open social data indicate human activities of a place, especially inside the built-up area. The use of both data sources is beneficial for urban land use type identification in urban areas. In addition, the sensitivity of features adopted in the similarity assessment was tested. If we used only the POI or two biophysical features derived from the Landsat images (*i.e.,* NDVI, NDBI), the obtained accuracies of the urban land parcels (Level I) in the built-up region are 41% and 31%, respectively. However, when combined, the accuracy reaches 75%. Moreover, some function types that were not identified in previous studies have been classified in our result, such as service buildings, medical and public places [28,39]. Hence, the use of both features at the parcel level produced more detailed and accurate land use maps than studies using single source data [23,28].

Remote sensing shows its advantages in the mapping of land cover dynamics over an aerial area. Combined with open social data, it is promising to map detailed urban land use over a large metropolitan area, which is advantageous compared with methods that focus only on discerning a specific land use type in a relatively small area [19,27]. Our approach is a prototype to serve this purpose. The impervious surface of the land cover map separated the urban area into built-up and non-built-up regions, so different classification strategies were used in terms of the distinctive characteristics of the two parts. Our experimental results indicate that both the biophysical attributes and the social economic features are crucial to urban land use type identification. For instance, community parcels are usually built with some vegetation coverage due to the consideration of eco-friendly human habitation. We can identify land use function with the dominant land cover types in the non-built-up region because minimal social data are considered in this area.

Although our approach can generate a detailed land use map over a large area, some limitations, such as confusable parcels with mixed functional types (e.g., the accuracy performance of institutional type in Table 3 is an example), should be considered. Furthermore, for most Level II urban land use categories, the accuracies are relatively lower, *i.e.,* the producer's accuracies for service buildings,

medical lands and administrative departments are 57%, 44% and 33%, respectively. Possible reasons may be that (1) parcels generated by road networks performed well in the urban core area but faced problems in rural areas because the road networks become sparse [40] and that (2) the POIs and their densities do not contain adequate detail to differentiate urban land uses; particularly for Level II classes. Therefore, the limited representativeness of parcels and lower POIs densities [16] in the rural areas may lead to a decline in accuracy based on the uncertainty analysis.

6. Conclusions

High-resolution land use maps are needed for academic research and urban management. However, complex and heterogeneous urban landscapes pose challenges to land use mapping. The use of both physical features (*i.e.,* spectral information) derived from remotely sensed data and social attributes (e.g., socio-economic function) derived from open social data can help to delineate the detailed land use patterns in urban areas. We developed an approach to combine the strength of these two types of data to identify land use types quickly over a large area and tested the effectiveness in Beijing. The use of both biophysical features and socioeconomic features resulted in a land use map with higher accuracy and more detail. The overall accuracy of the land use map for this extremely heterogeneous urban area reached 81.04% and 69.89% for the Level I and Level II categories, respectively. This approach can be applied to derive land use maps of a large area in a relatively short time wherever satellite data and open social data are available, especially for fast-growing urban areas in developing countries.

More efforts can be undertaken for further improvement. First, incorporating both road networks derived from open social data and segmentation derived from biophysical features (e.g., bands in Landsat images) may be helpful to generate more detailed parcels, especially for suburban areas in which road networks are relatively sparse. Second, the shapes and sizes of parcels can be considered when assigning possible land use types (e.g., parcels in industrial areas are relatively large, but parcels in residential areas are small). Third, non-linear approaches, such as neural networks, can be considered when building the relationships between physical and socioeconomic features and training samples [56,57] if the number of pre-collected training parcels is adequate.

Author Contributions: T.H. designed the study, performed most of the analysis, and led the writing. J.Y., X.L., and P. G. contributed to the writing.

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

References

1. Wang, L.; Li, C.C.; Ying, Q.; Cheng, X.; Wang, X.Y.; Li, X.Y.; Hu, L.Y.; Liang, L.; Yu, L.; Huang, H.B.; *et al*. China's urban expansion from 1990 to 2010 determined with satellite remote sensing. *Chin. Sci. Bull.* **2012**, *57*, 2802–2812. [CrossRef]
2. United Nations. *World Urbanization Prospects: The 2014 Revision*; United Nations: New York, NY, USA, 2015.
3. Li, X.; Yeh, A.G.-O. Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *Int. J. Geogr. Inf. Sci.* **2000**, *14*, 131–152. [CrossRef]
4. Li, X.; Liu, X.; Gong, P. Integrating ensemble-urban cellular automata model with an uncertainty map to improve the performance of a single model. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 762–785. [CrossRef]
5. Li, X.; Liu, X.; Yu, L. Aggregative model-based classifier ensemble for improving land-use/cover classification of Landsat TM Images. *Int. J. Remote Sens.* **2014**, *35*, 1481–1495. [CrossRef]
6. Gong, P.; Liang, S.; Carlton, E.J.; Jiang, Q.; Wu, J.; Wang, L. Urbanisation and health in China. *Lancet* **2012**, *379*, 843–852. [CrossRef]

7.　Lu, D.; Tian, H.; Zhou, G.; Ge, H. Regional mapping of human settlements in southeastern china with multi sensor remotely sensed data. *Remote Sens. Environ.* **2008**, *112*, 3668–3679. [CrossRef]

8.　Bateman, I.J.; Harwood, A.R.; Mace, G.M.; Watson, R.T.; Abson, D.J.; Andrews, B.; Binner, A.; Crowe, A.; Day, B.H.; Dugdale, S.; *et al.* Bringing ecosystem services into economic decision-making: Land use in the United Kingdom. *Science* **2013**, *341*, 45–50. [PubMed]

9.　Li, G.; Fang, C. Global mapping and estimation of ecosystem services values and gross domestic product: A spatially explicit integration of national "green GDP" accounting. *Ecol. Indic.* **2014**, *46*, 293–314. [CrossRef]

10.　Li, S.-M. Housing consumption in urban china: A comparative study of Beijing and Guangzhou. *Environ. Plan. A* **2000**, *32*, 1115–1134. [CrossRef]

11.　Huang, Y.; Jiang, L. Housing inequality in transitional Beijing. *Int. J. Urban Reg. Res.* **2009**, *33*, 936–956. [CrossRef]

12.　Shin, H.B. Residential redevelopment and the entrepreneurial local state: The implications of Beijing's shifting emphasis on urban redevelopment policies. *Urban Stud.* **2009**, *46*, 2815–2839. [CrossRef]

13.　Zhao, P.J.; Lu, B. Transportation implications of metropolitan spatial planning in mega-city Beijing. *Int. Dev. Plan. Rev.* **2009**, *31*, 235–261. [CrossRef]

14.　Xue, C.; Zheng, X.; Zhang, B.; Yuan, Z. Evolution of a multidimensional architectural landscape under urban regeneration: A case study of Jinan, China. *Ecol. Indic.* **2015**, *55*, 12–22. [CrossRef]

15.　Manonmani, R.; Suganya, G.M. Remote sensing and GIS application in change detection study in urban zone using multi temporal satellite. *Int. J. Geomat. Geosci.* **2010**, *1*, 60–65.

16.　Falcone, J.A. *U.S. Conterminous Wall-To-Wall Anthropogenic Land Use Trends (NWALT), 1974–2012*; USGS: Reston, VA, USA, 2015. Available online: http://dx.doi.org/10.3133/ds948 (accessed on 31 January 2016).

17.　Theobald, D.M. Development and applications of a comprehensive land use classification and map for the US. *PLoS ONE* **2014**, *9*, e94628.

18.　Donnay, J.; Unwin, D. Modeling geographical distributions in urban areas. In *Remote Sensing and Urban Analysis*; Donnay, J., Barnsley, M.J., Longley, P.A., Eds.; Taylor & Francis: New York, NY, USA, 2001; pp. 205–224.

19.　Wu, S.-S.; Qiu, X.; Usery, E.L.; Wang, L. Using geometrical, textural, and contextual information of land parcels for classification of detailed urban land use. *Ann. Assoc. Am. Geogr.* **2008**, *99*, 76–98. [CrossRef]

20.　Lu, D.; Weng, Q. Use of Impervious Surface in Urban Land Use Classification. *Remote Sens. Environ.* **2006**, *102*, 146–160. [CrossRef]

21.　Gong, P.; Howarth, P.J. The use of structural information for improving land-cover classification accuracies at the rural-urban fringe. *Photogramm. Eng. Remote Sens.* **1990**, *56*, 67–73.

22.　Shaban, M.A.; Dikshit, O. Improvement of Classification in Urban Areas by the Use of Textural Features: The Case Study of Lucknow City, Uttar Pradesh. *Int. J. Remote Sens.* **2001**, *22*, 565–593. [CrossRef]

23.　Hu, S.; Wang, L. Automated urban land use classification with remote sensing. *Int. J. Remote Sens.* **2012**, *34*, 790–803. [CrossRef]

24.　Wu, S.; Silvan, C.; Wang, L. Per-field urban land use classification based on tax parcel boundaries. *Int. J. Remote Sens.* **2007**, *28*, 2777–2801. [CrossRef]

25.　Liu, Y.; Liu, X.; Gao, S.; Gong, L.; Kang, C.; Zhi, Y.; Chi, G.; Shi, L. Social sensing: A new approach to understanding our socioeconomic environments. *Ann. Assoc. Am. Geogr.* **2015**, *105*, 512–530. [CrossRef]

26.　Ratti, C.; Frenchman, D.; Pulselli, R.M.; Williams, S. Mobile landscapes: Using location data from cell phones for urban analysis. *Environ. Plan. B: Plan. Des.* **2006**, *33*, 727–748. [CrossRef]

27.　Toole, J.L.; Ulm, M.; González, M.C.; Bauer, D. Inferring land use from mobile phone activity. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*; ACM: Beijing, China, 2012; pp. 1–8.

28.　Pei, T.; Sobolevsky, S.; Ratti, C.; Shaw, S.-L.; Li, T.; Zhou, C. A new insight into land use classification based on aggregated mobile phone data. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1988–2007. [CrossRef]

29.　Guande, Q.; Xiaolong, L.; Shijian, L.; Gang, P.; Zonghui, W.; Daqing, Z. Measuring social functions of city regions from large-scale taxi behaviors. In Proceedings of the 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), Seattle, WA, USA, 21–25 March 2011; pp. 384–388.

30.　Liu, Y.; Wang, F.; Xiao, Y.; Gao, S. Urban land uses and traffic "source-sink areas": Evidence from GPS-enabled taxi data in Shanghai. *Landsc. Urban Plan.* **2012**, *106*, 73–87. [CrossRef]

31. Gong, Y.; Liu, Y.; Lin, Y.; Yang, J.; Duan, Z.; Li, G. Exploring spatiotemporal characteristics of intra-urban trips using metro smartcard records. In Proceedings of the 2012 20th International Conference on Geoinformatics (GEOINFORMATICS), Hong Kong, China, 15–17 June 2012; pp. 1–7.

32. Sevtsuk, A.; Ratti, C. Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks. *J. Urban Technol.* **2010**, *17*, 41–60. [CrossRef]

33. Goodchild, M.F.; Hill, L.L. Introduction to digital gazetteer research. *Int. J. Geogr. Inf. Sci.* **2008**, *22*, 1039–1044. [CrossRef]

34. Crutcher, M.; Zook, M. Placemarks and waterlines: Racialized cyberscapes in post-Katrina Google Earth. *GeoForum* **2009**, *40*, 523–534. [CrossRef]

35. Elwood, S.; Goodchild, M.F.; Sui, D.Z. Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Ann. Assoc. Am. Geogr.* **2011**, *102*, 571–590. [CrossRef]

36. McKenzie, G.; Janowicz, K.; Adams, B. A weighted multi-attribute method for matching user-generated Points of Interest. *Cartogr. Geogr. Inf. Sci.* **2014**, *41*, 125–137. [CrossRef]

37. Wang, J.; Li, C.C.; Hu, L.Y.; Zhao, Y.Y.; Huang, H.B.; Gong, P. Seasonal land cover dynamics in Beijing derived from Landsat 8 data using a spatio-temporal contextual approach. *Remote Sens.* **2015**, *7*, 865–881. [CrossRef]

38. Long, Y.; Liu, X.J. Automated identification and characterization of parcels (AICP) with OpenStreetMap and Points of Interest. *Cornell University Library.* **2013**.

39. Erol, H.; Akdeniz, F. A per-field classification method based on mixture distribution models and an application to Landsat Thematic Mapper data. *Int. J. Remote Sens.* **2005**, *26*, 1229–1244. [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, New York, NY, USA, 12 August 2012; pp. 186–194.

41. Ministry of housing and urban-rural development (MoHURD). *Code for Design of Urban Road Engineering CJJ37.Beijing*; China Architecture & Building Press: Beijing, China, 2012.

42. Ministry of housing and urban-rural development (MoHURD). *The People's Republic of China on Urban and Rural Planning*; China Legal Publishing House: Beijing, China, 2007.

43. Xie, Y.T. The planning control of non-urban development land. *Urban Plan. Forum* **2005**, *158*, 35–41.

44. Ministry of housing and urban-rural development (MoHURD). *Current Land Use Classification*; China National Standardization Management Committee: Beijing, China, 2007.

45. Xian, G.; Crane, M. Assessments of urban growth in the Tampa Bay watershed using remote sensing data. *Remote Sens. Environ.* **2005**, *97*, 203–215. [CrossRef]

46. Aguilera, F.; Valenzuela, L.M.; Botequilha-Leitão, A. Landscape metrics in the analysis of urban land use patterns: A case study in a Spanish metropolitan area. *Landsc. Urban Plan.* **2011**, *99*, 226–238. [CrossRef]

47. Gong, P.; Wang, J.; Yu, L.; Zhao, Y.; Zhao, Y.; Liang, L.; Niu, Z.; Huang, X.; Fu, H.; Liu, S.; *et al.* Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens.* **2013**, *34*, 2607–2654. [CrossRef]

48. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chapman and Hall: New York, NY, USA, 1986; Volume 26, p. 76.

49. Li, X.; Gong, P.; Liang, L. A 30-year (1984–2013) record of annual urban dynamics of Beijing city derived from Landsat data. *Remote Sens. Environ.* **2015**, *166*, 78–90. [CrossRef]

50. Roy, D.P.; Wulder, M.A.; Loveland, T.R.; C.E, W.; Allen, R.G.; Anderson, M.C.; Helder, D.; Irons, J.R.; Johnson, D.M.; Kennedy, R.; *et al.* Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens. Environ.* **2014**, *145*, 154–172. [CrossRef]

51. Zha, Y.; Gao, J.; Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594. [CrossRef]

52. Gao, F.; de Colstoun, E.B.; Ma, R.; Weng, Q.; Masek, J.G.; Chen, J.; Pan, Y.; Song, C. Mapping impervious surface expansion using medium-resolution satellite image time series: A case study in the Yangtze river delta, China. *Int. J. Remote Sens.* **2012**, *33*, 7609–7628. [CrossRef]

53. Liu, X.; Hu, G.; Ai, B.; Li, X.; Shi, Q. A Normalized Urban Areas Composite Index (NUACI) based on combination of DMSP-OLS and MODIS for mapping impervious surface area. *Remote Sens.* **2015**, *7*, 17168–17189. [CrossRef]

54. Rozenstein, O.; Karnieli, A. Comparison of methods for land-use classification incorporating remote sensing and GIS inputs. *Appl. Geogr.* **2011**, *31*, 533–544.

55. Kuang, W. Spatio-temporal patterns of intra-urban land use change in Beijing, China between 1984 and 2008. *Chin. Geogr. Sci.* **2012**, *22*, 210–220. [CrossRef]

56. Karlsson, C. Clusters, Functional Regions and Cluster Policies, JIBS and CESIS Electronic Working Paper Series, CESIS, Stockholm, Sweden, 2007.

57. Wu, S.S.; Xu, B.; Wang, L. Urban land use classification using variogram-based analysis with an aerial photograph. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 813–822. [CrossRef]