Sensors and Materials, Vol. 35, No. 1 (2023) 269–284
MYU Tokyo

S & M 3167

Using Point of Interest Data and Satellite Observation for Urban Functional Zone Mapping

Meizi Yang,[1] Shisong Cao,[2] Hengrui Zhang,[1] Shuang Wu,[3] and Dayu Zhang[1*]

[1]School of Architecture and Urban Planning, Beijing University of Civil Engineering and Architecture, 1 Zhanlanguan Road, Xicheng District, Beijing 100044, China
[2]School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, 15 Yongyuan Road, Huangcun, Daxing District, Beijing 102612, China
[3]Beijing Institute of Surveying and Mapping, 60 Nanlishi Road, Xicheng District, Beijing 100044, China

(Received October 31, 2022; accepted January 11, 2023)

Keywords: point of interest, land function, TF-IDF, Luojia-1, Spearman rank correlation coefficient

Today, China's development mode has shifted from extensive management to meticulous management, so it is inevitable that data will be used to conduct quantitative analysis. In this study, we synthesized multi-source data and the term frequency-inverse document frequency (TF-IDF) algorithm to determine the dominant function of land within the Fifth Ring Road of Beijing, China. Additionally, we explored how point of interest (POI) distributions affect nighttime artificial light radiation using the Spearman correlation coefficient. The overall accuracy of land functional classification in the central urban area of Beijing was 88.07%, and the kappa coefficient was 85.28%. The core area of Beijing is dominated by commercial and business facilities, accounting for 29.37% of the total area, followed by residential (22.02%), mixed-use (16.51%), green space (10.62%), administration and public service (9.66%), transportation (9.31%), and industrial (2.52%) areas. The number of commercial and business facilities, administration and public services, and residential areas correlate positively with the vegetation adjusted nighttime light (NTL) urban index (VANUI) value. In contrast, the number of industrial facilities yielded a negative correlation with VANUI. This study provides support for the optimization of the functional layout and the distribution of public resources in the central urban area of Beijing.

1. Introduction

With the rapid development of the social economy and expansion, China's urbanization has considerably been enhanced in recent decades. The rate of urbanization has climbed from 36% in 2000 to 64% in 2021, and China has become the second-largest economy in the world. However, increasing urbanization has also brought many problems, such as housing tension, traffic congestion, the lack of public space, and uneven distribution of resources. These contradictions have had a negative effect on construction to make Beijing a livable city and sustainable development.(1) Meanwhile, existing studies show that urban land function has a

*Corresponding author: e-mail: zhangdy@bucea.edu.cn
https://doi.org/10.18494/SAM4213

ISSN 0914-4935
https://myukk.org/

strong correlation with urban environments, vitalities, and equities.[2–4] Therefore, a better investigation regarding urban land function identification, as well as how urban land function affects urban planning, urban design, and urban management, is required.

Generally speaking, identifying urban land functions includes field investigation by professionals, construction permits, maps, and planning documents. However, these methods require extensive time and manual work. Moreover, such methods are highly subjective and have low accuracy.[5] With the rapid development of remote sensing technology, satellite observation can achieve accurate and detailed information on land cover.[6–8] It is noteworthy that it is hard to directly identify land functions, i.e., commercial and business zones, from satellite images without additional social sensing data. Fortunately, with the advent of the information and data era, using social media, mobile signaling, location-based services (LBS), and point of interest (POI) data have shown great potential to identify urban land functions.[5,9,10] Among these methods, POI data are the most widely used owing to the advantages of detailed information coverage, high accuracy, frequent updates, and easy access.[11] Additionally, existing literature demonstrates that POI data can clearly illustrate human activities and social economy; with this in mind, POI data are recognized as a suitable means for land function identification.[2,12,13]

There are several problems using POI data to explore land functions. First, OpenStreetMap (OSM) data, which is often utilized to divide land parcels, is regarded as a basic geography unit for functional zone measurements.[14] However, when employing OSM data, the spatial scale of land parcels is coarse, resulting in low measurement accuracy. For example, many mixed-use functional zones show up in the results using coarse spatial units (e.g., land parcels), which is inconsistent with reality. Second, different impacts of different categories concerning POI data with respect to the land function have been observed.[2,5] When identifying urban land functions, it is necessary to assign varying weight coefficients to different POIs. Finally, many studies explore the distributions of land functions, but few studies consider the correlation between urban functions and other urban development indicators, such as thermal environments, economic development, and vitality.

On the basis of this discussion, we chose the core area of the city of Beijing, China, as the study area and employed POI data and the term frequency-inverse document frequency (TF-IDF) algorithm to determine POI weights; we also used road data from OSM overlay fishnets to identify land function at 300 m of spatial resolution. In addition, we used the Spearman correlation model to analyze the correlation between land function and vegetation adjusted nighttime light (NTL) urban index (VANUI). Doing so provides a basis for more reasonable urban resource allocation and land function replacement.

2. Study Area and Data

2.1 Study area

Beijing lies between longitude 115.7°-117.4°E and latitude 39.4°-41.6°N, with its center at 39° 54′ 20″ N and 116° 25′ 29″ E. As the capital of China and one of the world's top ten international metropolises, the city has a total area of 16410 km^2. By 2021, the permanent population and

annual GDP reached 21.886 million and 4026.96 billion yuan (RMB), respectively. Beijing is at the forefront of the world in terms of construction scale, population, and economic development. However, because of the rapid economic development and high concentration of the population, many problems have been exposed, like the inconsistency of land function and urban development level, the imbalance in the distribution of resources, and other issues that significantly constrain the sustainable development of the city. Considering the quantity and accuracy of POI, we selected the area within the Fifth Ring Road of Beijing as the study area.

2.2 Data and data preprocessing

2.2.1 OSM data

Scholars favor OSM data because of continued improvements in its quality and its free availability. It provides water systems, buildings, road networks, and POI data. Among these, the road network has better accuracy and integrity, and thus can be widely used to determine the basic units (parcels) in the land function classification. We used road network data downloaded from the OSM website to divide the basic spatial units. Before this, we deleted irrelevant information such as footways, raceways, cycleways, and platforms. We reserved motorways, primary roads, secondary roads, tertiary roads, residential roads, living streets, and service roads. Broken lines, duplicate lines, and incorrect lines were deleted. According to the road grade, a buffer zone method was developed to obtain the final road networks. A comparison between road networks before and after preprocessing is shown in Fig. 1.

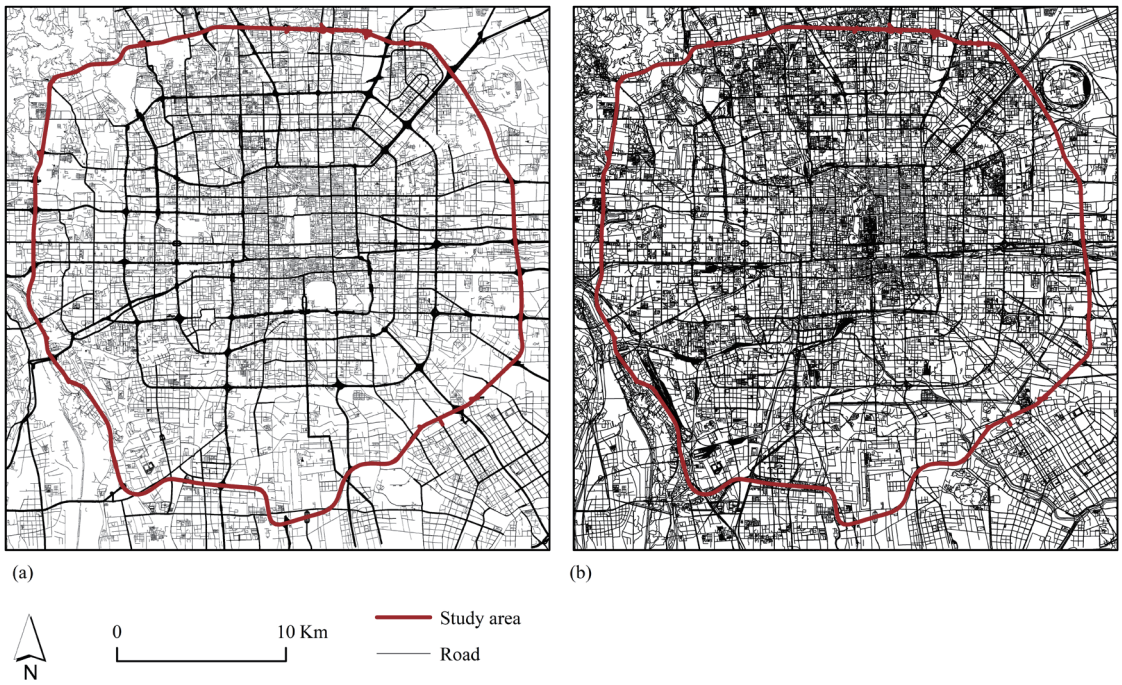

Fig. 1. (Color online) A comparison between road networks before and after removing irrelevant information: (a) after preprocessing and (b) before preprocessing.

2.2.2 POI data

The POI data was acquired from the Baidu Map Open Platform. First, Python programing language was used to generate the POI data, and a total of 963374 pieces of records were obtained. The data were then sorted, and 258042 valid POIs were acquired. According to urban land use and planning classification standards (GB 50137-2011) released by the Ministry of Housing and Urban-Rural Development of the People's Republic of China, we reclassified POI records into six categories, including administrative and public services (A), commercial and business facilities (C), green space (G), industrial (M), residential (R), and streets and transportation (S). Details of POI data categories are shown in Table 1.

2.2.3 Remotely sensed data

Nighttime light (NTL) data is closely related to human activities, economic development, urban vitality, and heat pollution in cities, and thus the data is widely used in investigating urban issues.[15] In this study, we used Luojia-1 NTL and POI data to conduct a correlation analysis and explored the interactions between urban functions and human activities. The Luojia-1 satellite is the world's first professional artificial light-detection satellite; it carries a high-sensitivity night light sensor with a spatial resolution of 130 m. Notably, the spatial resolution of Luojia-1 NTL data is considerably higher than that of the widely used Defense Meteorological Satellite Program - Operational Line-Scan System (DMSP-OLS) and the National Polar-Orbiting Partnership - Visible Infrared Imaging Radiometer Suite (NPP-VIIRS) night light data. Thus, it can provide support for urban research at the mesoscopic and microscopic scales. However, oversaturation phenomena in the nighttime light data of the Luojia-1 satellite restrict the representation of spatial structure changes in the urban area. We used the normalized difference vegetation index (NDVI) from Landsat 8 images associated with NTL data to generate the VANUI to reduce the influence of NTL oversaturation.[16] The equations for VANUI are expressed as:

Table 1
POI data classification system.

Urban function category	Description	Number	Proportion (%)
Residential (R)	Residential and residential-related facilities.	21098	8.18
Commercial and Business Facility (C)	Catering service, shopping service; accommodation service, financial and insurance services, and life service.	156498	60.65
Administration and Public Service (A)	Science, education, and culture, sports and leisure, medical care, government agencies, and social organizations.	76061	29.48
Industrial (M)	Corporation.	990	0.38
Green Space (G)	Park and plaza, and scenic spots.	3354	1.30
Street and Transportation (S)	Transportation service facilities.	41	0.02

$$NTR_{nor} = \frac{NTL - NTL_{min}}{NTL_{max} - NTL_{min}} \tag{1}$$

$$NDVI_{max} = MAX[NDVI_1, NDVI_2, \cdots, NDVI_2] \tag{2}$$

$$VANUI = (1 - NDVI_{max}) \times NTL_{nor} \tag{3}$$

where NTL_{nor} is the normalized NTL data; NTL_{max} and NTL_{min} are the maximum and minimum values in an NTL image, respectively; $NVDI_{max}$ is the maximum value in the multi-temporal composite NDVI data in 2018; $NDVI_1$, $NDVI_2$, …, $NDVI_n$ are the multitemporal NDVI images from Landsat 8 in 2018.

3. Methods

3.1 Spatial unit partitioning

City blocks, as the basic unit of urban function, can be generated using road networks. Most related studies used the road network for city block delineation. Note that the large scale and low precision of neighborhoods can lead to an increase in the number of mixed land functional zones, and certain small-scale functional zones cannot be identified.(5) Thus, on the basis of the preprocessed OSM road networks as the basic unit for neighborhood division, the parcels were further divided using a 300 m × 300 m fishnet to improve spatial resolution.

3.2 Urban land function identification

3.2.1 Single-use function identification

The TF-IDF statistical algorithm was used to calculate the weights of various types of POIs in the parcels. The TF-IDF model is a standard weighting method selected for information retrieval and data mining. The main idea is that, if a word or phrase appears in an article with high frequency and rarely appears in other articles, it can be considered to have good category differentiation ability and is suitable for classification. The specific formulas are

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}, \tag{4}$$

$$idf_i = \lg \frac{|D|}{\left|\left\{j : t_i \in d_j\right\}\right|}, \tag{5}$$

$$tfidf_{i,j} = tf_{i,j} \times idf_i, \tag{6}$$

where i denotes word, j denotes document, $n_{i,j}$ means the number of occurrences of the word i in the document j, and $\sum_k n_{i,j}$ represents the sum of all occurrences of words in the document j; $|D|$ refers to the total number of documents, and $|\{j: t_i \in d_j\}|$ is the total number of documents that contain the word i. In this study, documents represent parcels, and words are POIs in parcels.

The specific approach includes three key steps. First, the preprocessed POI data and the fishnet network of each parcel were spatially overlaid to obtain the number of different POI categories corresponding to each parcel. Second, the weights of each category were calculated on the basis of the TD-IDF algorithm. Third, the category with the maximum weight can be regarded as the dominant function for that parcel.

3.2.2 Mixed-use function identification

Following Huang *et al.*,[5] the Grubbs criterion was employed to identify the mixed-use land function. The ratio category index was calculated for each POI within a spatial unit:

$$CR_i = \frac{tfidf_{i,j}}{\sum_{i=1}^{n} tfidf_{i,j}} \times 100\%, (i = 1, 2, ..., n), \tag{7}$$

where CR_i is the proportion of the functional intensity of the ith POI type to the functional intensity of all POI types in a specific spatial unit. In addition, the Grubbs criterion was used to determine whether a certain type of POI yields a significantly higher weight than others in the parcel. If it does not appear, the parcel can be labeled as a mixed land use. The formula for the calculation is as follows:

$$G_{max} = \frac{CR_{max} - \overline{CR}}{S}, \tag{8}$$

where G_{max} is the maximum deviation value, CR_{max} is the maximal proportional value of the category of the ratio of the POI, $\overline{CR}$ is the average value of the proportional size of the category of the ratio of each POI type in the spatial unit, and S refers to the standard deviation of this dataset. The resulting offset value was compared with the Grubbs threshold, $G(n, p)$ (where n is the number of POI types within a certain parcel and p is the confidence probability. The value chosen in this study is 95%). If $G_{max} < G(n, p)$, the parcel can be labeled as a mixed land function.

3.3 Data analysis

3.3.1 Kernel density estimate

Kernel density analysis is a non-parametric estimation spatial partitioning method that uses the spatial properties of data to calculate the density of elements in their surrounding neighborhoods and then study the distribution characteristics of the spatial data.[17] In this study, the spatial analysis tool in ArcGIS 10.7 was used to compute kernel densities of various POIs within the fifth ring road of Beijing. The formula is

$$f_n(x) = \frac{1}{nh^2} \sum_{j=1}^{n} k\left(\frac{x - x_j}{h}\right), \tag{9}$$

where $f_n(x)$ is the kernel density value at the spatial location point x; h is the value range regarding thresholds; k is the default weight kernel function; n is the number of points in the analysis range; $(X—X_j)$ denotes the distance from the location x to the event j; $h > 0$ is the band, i.e., the search radius of the estimation of kernel width density; a larger value of $f_n(x)$ indicates a denser POI.

3.3.2 Spearman correlation coefficient

The Spearman rank correlation coefficient is a nonparametric measure of the dependence of two variables. It evaluates the correlation of two statistical variables using a monotonic equation.[18,19] The formula for the calculation is

$$P = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}, \tag{10}$$

where x_i is the rank of the ith data of the random variable X, y_i is the rank of the ith data of the random variable Y, and $\bar{x}$ and $\bar{y}$ denote the expectations of X and Y, respectively.

4. Results

4.1 Functional area identification results

4.1.1 Spatial distribution of different types of POI

Figure 2 provides the results of the kernel density determination for various POIs. As shown, the distribution of different functional zones in the city had a strong spatial heterogeneity due to the combined effects of historical development, planning guidance, and natural conditions.

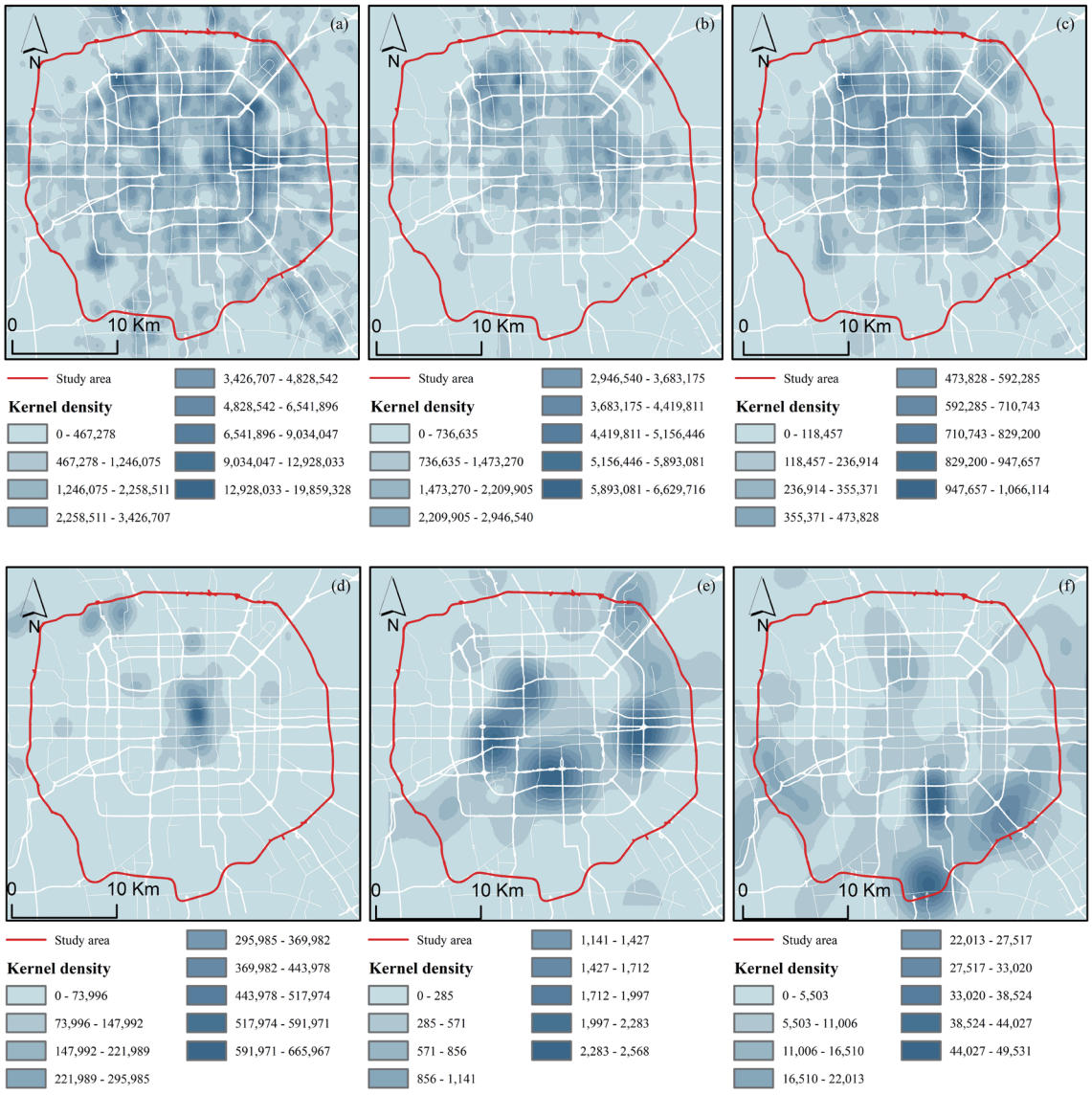

Fig. 2. (Color online) Results of the determination of kernel density for various POIs: (a) commercial and business facilities, (b) administration and public service, (c) residential, (d) green space, (e) street and transportation, and (f) industrial.

Commercial and business POIs were characterized by spreading throughout regions along the main traffic roads of the city, and clustering was observed near the Chaoyang central business district (CBD) areas, streets serving the financial activities in the Xicheng district, and Zhongguancun science park in the Haidian district. We found few differences in the number of commercial and business facilities in different regions, which were distributed relatively evenly, while an obvious difference was found in their types. For example, in the areas between the second and third ring roads, extensive commercial streets, i.e., Xidan and Wangfujing streets, were observed. However, many large commercial shopping centers and office buildings, i.e.,

Chaowai Soho, the world trade plaza, and Xinzhongguan, were found between the third and fourth ring roads. Meanwhile, in the places between the fourth and fifth ring roads, there were mostly wholesale markets and office buildings in industrial parks, e.g., Xinfadi wholesale market.

Administration and public service facilities were mostly distributed in the central and northern parts of the city. As can be seen, around the college road, many universities, i.e., Tsinghua University, Peking University, Beijing University of Geosciences, Beijing University of Aeronautics and Astronautics, and Beijing University of Science and Technology, were clustered, making this region a public service gathering area. Residential facilities were mainly concentrated between the second and fourth ring roads. As shown, the spatial distribution of residential facilities displayed a strong consistency with commercial and business facilities. Green space was concentrated within the second ring road, including the Forbidden City, Temple of Heaven, Longtan Lake Park, Taoranting Park, Beihai Park, and Shichahai Park. The rest of the green space facilities were concentrated in the northeastern parts of the city, e.g., Yuan Ming Yuan, the Summer Palace, and Olympic Forest Park. Most of the parks are closely related to Beijing's history and cultural environments.

Transportation facilities were mainly concentrated around railroad stations, e.g., Beijing West Station, Beijing North station, Beijing South Station, and Beijing Station. Industrial-related POIs were primarily located around the south fifth ring road, mainly in the Chaoyang and Fengtai districts. The southern part of Beijing was in the downwind direction. This location would significantly reduce the diffusion of industrial pollution, and it is well in line with the planning for the Chaoyang and Fengtai districts as the industrial base of Beijing.

4.1.2 Spatial distribution of land function

Figure 3 shows the results of the spatial division of functional zones. As shown, lots of low-end commercial and business facilities in the core area of the second ring road were evacuated to other regions because of the completion of several rounds of special actions of "evacuating, improving, and promoting." Additionally, many low-quality residences and industries were evacuated to other regions. The core area was observed to be an attractive place with facilities and parks. In contrast, land functions were more complex in the area between the second and fourth ring roads and all types of function were evenly distributed. Most of the blocks that were divided by the road network showed a circular pattern of green space-residential-mixed use-business and commercial facilities or administration and public service facilities-residential-mixed use-business and commercial facilities from the inside to the outside. In the areas between the fourth to fifth ring roads, there were massive differences in functions between the northern and southern parts of Beijing. The northern part concentrated on public service facilities, i.e., colleges and universities. At the same time, technology parks and residences were clustered around the colleges and universities. In addition, many commercial office facilities were clustered around Olympic Forest Park. By contrast, the southern part of Beijing was dominated by inefficient commercial and industrial functional zones.

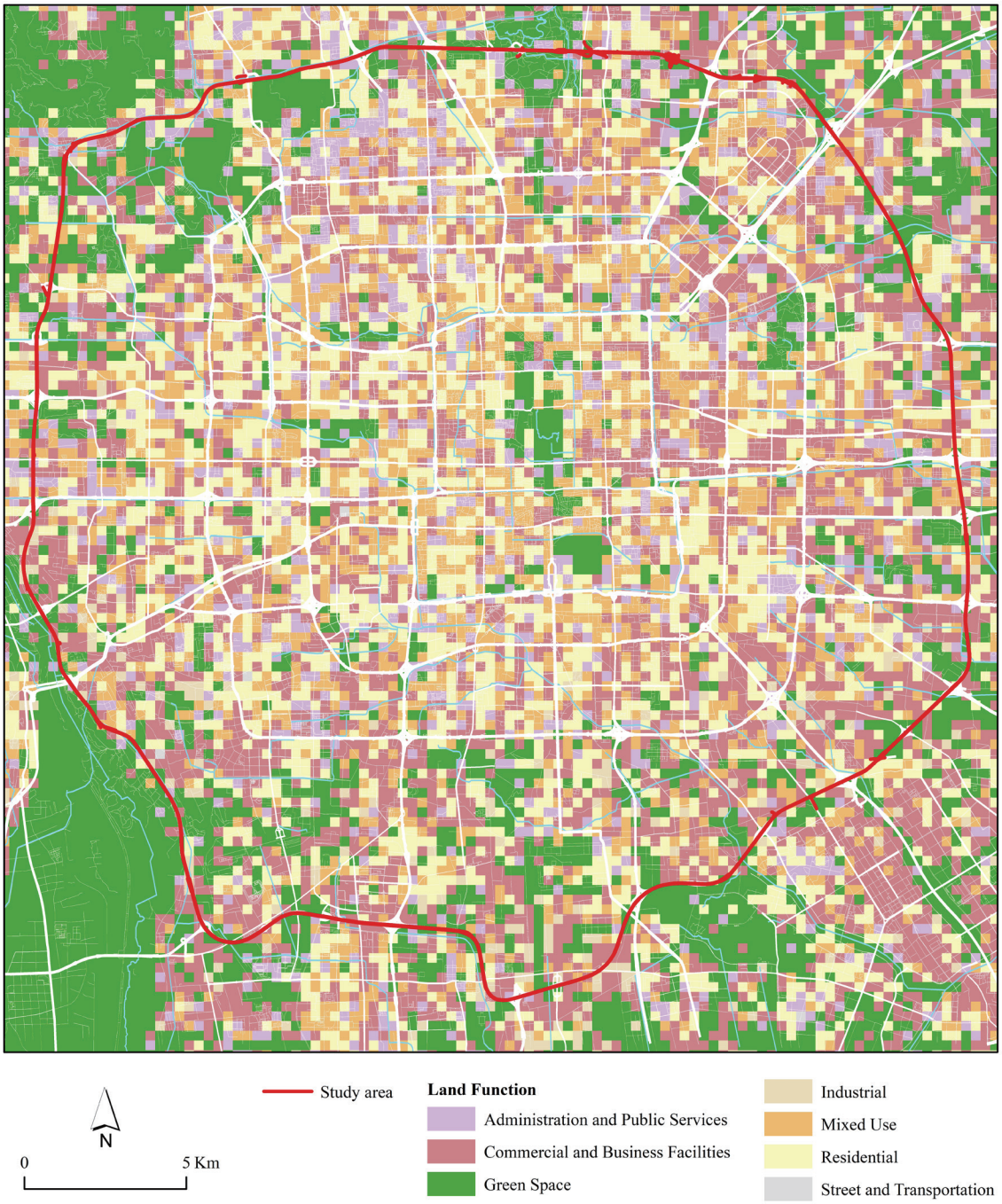

Fig. 3. (Color online) Result of urban land function identification.

Table 2 shows land areas and proportions for different functional zones. In the areas within the fifth ring road of Beijing, business and commercial (C) facilities occupied the highest land area and proportion (with an area of 295.91 km^2, accounting for 29.37% of land area), followed by residential (R) (with an area of 221.91 km^2, accounting for 22.02% of land area), mixed use

Table 2
Land areas and proportions of different functional zones.

Function zone	Area (km^2)	Proportion (%)
C	295.91	29.37
R	221.91	22.02
F	166.36	16.51
G	107.00	10.62
A	97.29	9.66
S	93.81	9.31
I	25.35	2.52

(F) (with an area of 166.36 km^2, accounting for 16.51% of land area), green space (G) (with an area of 107.00 km^2, accounting for 10.62% of land area), administration and public service facilities (A) (with an area of 97.29 km^2, accounting for 9.66% of land area), transportation (S) (with an area of 93.81 km^2, accounting for 9.31% of land area), and industrial (I) (with an area of 25.35 km^2, accounting for 2.52% of land area).

4.1.3 Spatial distribution of mixed-use land

Figure 4 shows that the distribution of mixed-use land functional zones was mainly located within the third ring road, indicating a decreasing trend from the center to the periphery. The areas with mixed-use function were concentrated in the Fangzhuang regions, Beijing West Station, and Xidan railway stations. Table 3 shows land areas and proportions for each mixed-use functional zone. The C/R/A (area = 87.77 km^2, accounting for 52.76% of land area) yielded the highest land area, and proportions in the mixed-use land function zones, followed by C/R/A/G (area = 29.72 km^2, accounting for 17.87% of land area), C/A (area = 19.89 km^2, accounting for 11.96% of land area), C/R/A/M (area = 8.74 km^2, accounting for 5.25% of land area), C/R (area = 2.96 km^2, accounting for 1.78% of land area), C/A/G (area = 2.96 km^2, accounting for 1.78% of land area), and C/R/M (area = 2.47 km^2, accounting for 1.48% of land area).

4.1.4 Accuracy assessment

To assess accuracy, we used the Baidu Map to choose some typical areas of different functional lands for qualitative comparison analysis. Yuyuantan Park, Longtan Lake Park, Beihai Park, and the Forbidden City were chosen to verify the accuracy of green spaces. In addition, Tsinghua University, Peking University, Beijing University of Civil Engineering and Architecture, the U.S. Embassy, and the Beijing Municipal Government Service Center were chosen to evaluate the accuracy of the classification of administration and public service. Jianwai Soho, Xidan, and Wangfujing were selected to assess the accuracy of business and commercial areas. Baiwanzhuang and Ganjiakou residential areas were selected for evaluating residential areas. Beijing West Railway Station, Beijing South Railway Station, and Beijing Railway Station were selected for the assessment of the accuracy of land devoted to transportation facilities. Notably, Beijing North Railway Station was chosen for mixed-use

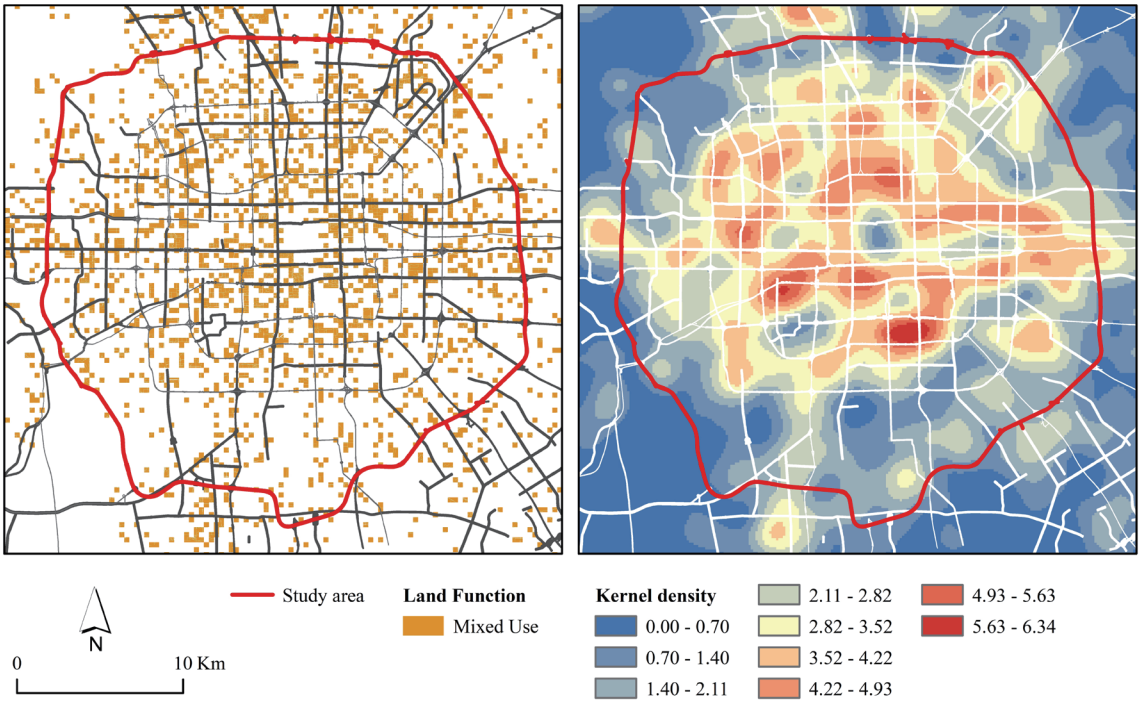

Fig. 4. (Color online) Spatial distribution of mixed-use land functional zones.

Table 3
Land areas and proportions for each mixed-use functional zone.

Function zone	Area (km^2)	Proportion (%)
C/R/A	87.77	52.76
C/R/A/G	29.72	17.87
C/A	19.89	11.96
C/R/A/M	8.74	5.25
C/R	3.26	1.96
C/A/G	2.96	1.78
C/R/M	2.47	1.48
Others	11.55	6.94

functional land evaluation. The southwest suburban food freezing plant and others were selected for the assessment of the accuracy of industrial functional zones. The results of land function classification in these typical areas are all consistent with the information provided by the Baidu Maps.

To derive more precision in the accuracy of the land functional classification results, the precision of the classification results was evaluated using the error matrix method, which has four main accuracy indicators: user accuracy, producer accuracy, overall accuracy, and kappa coefficient. In this study, 218 points in four areas (i.e., Zhanlan road area, Xiju area, college road area, and Jianwai Soho area) were selected for validation, and the results are shown in Table 4. High accuracy was observed using our model for urban functional zone mapping. The user

Table 4
Error matrix of urban functional land classification.

Reference data	Classes							
	C	A	R	M	G	S	F	PA
C	44	0	0	2	0	0	1	0.94
A	0	43	3	0	0	0	4	0.86
R	0	4	32	0	0	0	3	0.82
M	0	0	0	7	2	0	2	0.64
G	0	0	0	0	47	0	0	1.00
S	0	0	0	0	1	1	0	0.50
F	2	0	2	0	0	0	18	0.82
UA	0.96	0.91	0.86	0.78	0.94	1.00	0.64	

accuracy of mixed-used functional zones was relatively low, as was the production accuracy of transportation and industrial areas. However, both the production and user accuracy of the remaining land function types were above 80%, with an overall accuracy of 81.8%. The overall accuracy of the classification model proposed in this study reached 88.07%, and the kappa coefficient was 85.28%, which is highly consistent with the actual situation.

4.2 Correlation analysis

4.2.1 Spatial distribution of nighttime lights

Figure 5 shows the spatialized results of VANUI after resampling. As shown, the area with the strongest light intensity within the fifth Ring Road of Beijing was found to be located in the Jianguomen district. The area around Wangfujing and the CBD area from the East third to fourth ring roads had concentrated and very high brightness. It was found that these areas were mainly large commercial centers and close to Chang'an Street and the ring road.

4.2.2 Results of Spearman correlation coefficient

Table 5 shows the results using the Spearman correlation analysis of the number of each POI type and VANUI. When the absolute coefficient value lies between 0.8 and 1.0, the urban functional zones yielded a very strong correlation with VANUI values. A strong correlation between them can be identified when the absolute coefficient value lies between 0.6 and 0.8. A relatively strong correlation can be identified when the absolute coefficient value is between 0.4 and 0.6. However, a weak correlation can be identified when the absolute coefficient value is below 0.4.

Among varying POI types, A/C, R/C, and R/A showed strong positive correlations. M and VANUI had a very strong negative correlation. C and A were strongly positively correlated with VANUI. R and VANUI yielded a relatively strong positive correlation. The rest showed weak correlations with VANUI.

Fig. 5. (Color online) Spatial distribution of VANUI values.

Table 5
(Color online) Results of Spearman rank correlation coefficient.

	C	A	R	G	S	M	VANUI
C	1**	0.732**	0.693**	0.128**	0.041**	0.150**	0.753**
A		1**	0.706**	0.222**	0.025**	0.045**	0.747**
R			1**	0.145**	0.018*	0.055**	0.542**
G				1**	−0.006	−0.012**	0.201**
S					1**	0.006	−0.070
M						1**	−0.836**
VANUI							1**

5. Discussion

5.1 Spatial unit division for functional zone identification

Existing studies mainly used the road network as the spatial unit of functional land identification, yet this study superimposed a 300 m × 300 m grid on the area of interest for the delineation, hoping to improve the classification accuracy. Compared with the investigation conducted by Ning *et al.*,[20] high accuracy was observed in this study. However, there are many problems associated with using a simple grid. For instance, the size of the land parcels should not be the same in different areas; some buildings are also separated by the grid. These issues should be addressed in future research.

5.2 Identification of transportation function

In this study, only a very small amount of valid POI data about traffic was obtained, thus reducing its accuracy. Meanwhile, with the transit-oriented development (TOD) of cities, most of the transportation sites have been compounded for use. For example, Beijing North station has integrated subway, railroad, commercial, and office areas by means of three-dimensional development. In this context, such transportation areas would be divided into mixed-use land, thus leading to an underestimation of the transportation functional zone area.

5.3 Interpretation of NTL data

It is noteworthy that we innovatively linked the POI and night light remotely sensed data and measured the varying correlations between urban functions and VANUI values. However, for urban planning and design, the urban information that can be revealed within the NTL is more critical. Thus, in future research, the NTL remotely sensed data should be used as a carrier of urban information to explore the consequence of urban land functions.

6. Conclusions

In this study, we used OSM road network data and Baidu POI data to estimate the functional distribution within the fifth ring road of Beijing using the TD-IDF algorithm and proposed an urban functional land classification model based on POI. Through qualitative comparison and analysis with some representative areas in Beijing, which provide clear land functional types, we found that the classification results effectively reflect the actual land functional situation. In addition, with the assistance of field investigation and the Baidu map, the quantitative accuracy was verified and evaluated by the error matrix method. Upon examination, the overall accuracy of the proposed model reached 88.07%, with a kappa coefficient of 85.28%. This value suggests that the classification results are highly consistent with the actual situation. Finally, we explored the correlations between the number of POIs and VANUI using the Spearman correlation model.

Notably, the method of urban land function classification proposed in this study makes it easier to acquire data and to show a more straightforward procedure than existing methods such as cell phone signaling and remote sensing interpretation. This method can provide the distribution of land functions under a complex situation in a concise manner. Using roads overlaid with fishnets improves the measurement accuracy compared with the simple division using single road networks. In this study, we attempted to find interactions between urban land functions and human activities, which can provide technical support for the site selection of various facilities, transportation land planning, as well as the construction of a fifteen-minute living circle, and thus provide a reference for future study regarding the division of urban functions and urban spatial structure classification.

Acknowledgments

This work was supported by the (1) Categorical Development Quota Project - Doctoral Student Research Capacity Enhancement Project (2022), No. 31081022002 and (2) Beijing Key Laboratory of Urban Spatial Information Engineering, No. 20220103.

References

1 C. Gu: Sci. Chin. Earth Sci. **62** (2019) 1351. https://doi.org/10.1007/s11430-018-9359-y
2 W. Huang, L. Cui, M. Chen, D. Zhang, and Y. Yao: Int. J. Geog. Inf. Sci. **36** (2022) 1905. https://doi.org/10.1080/13658816.2022.2040510
3 J. Lee, J. Arts, F. Vanclay, and J. Ward: Sustainability **12** (2020) 5907. https://doi.org/10.3390/su12155907
4 N. Jogan, F. Kuzmic, and U. Silc: Urban Ecosyst. **25** (2022) 149. https://doi.org/10.1007/s11252-021-01140-4
5 C. Huang, C. Xiao, and L. Rong: Remote Sens. **14** (2022) 4201. https://doi.org/10.3390/rs14174201
6 S. Hu and L. Wang: Int. J. Remote Sens. **34** (2013) 790. https://doi.org/10.1080/01431161.2012.714510
7 C. Fu, X. Song, and K. Stewart: Remote Sens. **11** (2019) 2965. https://doi.org/10.3390/rs11242965
8 Y. Zhang, K. Qin, Q. Bi, W Cui, and G. Li: Remote Sens. **12** (2020) 1831. https://doi.org/10.3390/rs12111831
9 J. J. Arsanjani, M. Helbich, M. Bakillah, J. Hagenauer, and A. Zipf: Int. J. Geog. Inf. Sci. **27** (2013) 2264. https://doi.org/10.1080/13658816.2013.800871
10 W. Tu, J. Cao, Y. Yue, S. Shaw, M. Zhou, Z. Wang, X. Chang, Y. Xu, and Q. Li: Int. J. Geog. Inf. Sci. **31** (2017) 2331. https://doi.org/10.1080/13658816.2017.1356464
11 A. Crooks, D. Pfoser, A. Jenkins, A. Croitoru, A. Stefanidis, D. Smith, S. Karagiorgou, A. Efentakis, and G. Lamprianidis: Int. J. Geog. Inf. Sci. **29** (2015) 720. https://doi.org/10.1080/13658816.2014.977905
12 K. Janowicz: Trans. GIS **16** (2012) 351. https://doi.org/10.1111/j.1467-9671.2012.01342.x
13 A. Lin, X. Sun, H. Wu, W. Luo, D. Wang, D. Zhong, Z. Wang, L. Zhao, and J. Zhu: IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **14** (2021) 8864. https://doi.org/10.1109/JSTARS.2021.3107543
14 Z. Wang, D. Ma, D. Sun, and J. Zhang: PLoS One **16** (2021) e0251988. https://doi.org/10.1371/journal.pone.0251988
15 X. Li, X. Li, D. Li, X. He, and M. Jendryke: Remote Sens. Lett. **10** (2019) 526. http://doi.org/10.1080/2150704X.2019.1577573
16 J. Ou, X. Liu, P. Liu, and X. Liu: Int. J. Appl. Earth Obs. Geoinf. **81** (2019) 1. https://doi.org/10.1016/j.jag.2019.04.017
17 D. Marin, M. Tang, I. B. Ayed, and Y. Boykov: IEEE Trans. Pattern Anal. Mach. Intell. **41** (2019) 136. https://doi.org/10.1109/TPAMI.2017.2780166
18 H. Y. Song and S. Park: KSII Trans. Internet Inf. Syst. **14** (2020) 1951. https://doi.org/10.3837/tiis.2020.05.005
19 E. Van den Heuvel and Z. Zhan: Am. Stat. **76** (2022) 44. https://doi.org/10.1080/00031305.2021.2004922
20 X. Ning, Y. Liu, H. Wang, M. Hao, B. Yang, and M. Wang: Geogr. Geo-Inf. Sci. **34** (2018) 42. https://doi.org/10.3969/j.issn.1672-0504.2018.06.007