Remote Sensing of Environment 202 (2017) 210–221

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Using multi-source geospatial big data to identify the structure of polycentric cities

Jixuan Cai [a], Bo Huang [a,b,*], Yimeng Song [a,*]

[a] Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong
[b] Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 19 July 2016
Received in revised form 11 June 2017
Accepted 28 June 2017
Available online 6 July 2017

Keywords:
Nighttime light image
Social media
Image segmentation
Spatial statistics
Polycentric structure
Subcenter

ABSTRACT

Identifying the structure of a polycentric city is vital to various studies, such as urban sprawl and population movement dynamics. This paper presents an efficient and reliable method that uses multi-source geospatial big data, including nighttime light imagery and social media check-in maps, to locate the main center and subcenters of a polycentric city. Unlike traditional methods that rely on statistical data categorized by administrative units, the proposed method can effectively identify the boundaries of urban centers, and the data source guarantees a timely monitoring and update. Four main procedures are involved: 1) a new observation unit is developed using object-oriented segmentation; 2) main centers are located using cluster analysis (Local Moran's I); 3) subcenter candidates are selected using significant positive residuals from geographically weighted regression (GWR); and 4) final centers are filtered using global natural breaks classification (NBC). These steps can be reproduced in different regions. To evaluate the effectiveness, the method was applied to three rapidly developing Chinese cities: Beijing, Shanghai, and Chongqing with different natural and economic characteristics. The performance of the proposed method has been carefully evaluated with qualitative and quantitative analyses. Comparative experiments were also conducted across different datasets to prove the benefits of combining a social media check-in map with remotely sensed imagery in a human environment study.

1. Introduction

Urbanization has led to an increase in the number of urban dwellers and significant changes to urban structures (Bai et al., 2014). In recent decades, more polycentric cities have emerged, resulting from the previously close-by but independent urban settlements that become a larger and more integrated city-system (Liu and Wang, 2016). The studies of polycentric city structure have been undertaken at different geographical scales, including the inter-city scale and the intra-city scale (Yang et al., 2015). A polycentric city at the inter-city scale usually covers more than one urban areas, as well as satellite cities, towns and intervening rural areas that are socio-economically tied to the urban core (Liu and Wang, 2016).

The components of the intra-city polycentric structure usually include the main center and the subcenters (McMillen and McDonald, 1997). The main center is the core of a city and generally covers the central business district (CBD). Subcenters are areas with greater densities of human activity than nearby locations within a city, which include edge cities and satellite towns. Such areas enjoy the benefits of agglomeration, but offer lower commuting costs for citizens and cheaper land costs to corporations than the urban downtown (McMillen, 2001). Accurately delineating the polycentric structure of a city is important for a better understanding of urban expansion and provides the public and city managers with information needed to evaluate the effectiveness of planning layouts. However, the distribution of urban centers is influenced by a variety of topographic and socio-economic circumstances, which are rarely parameterized into circles or ellipses (Redfearn, 2007).

Our knowledge of the urban structure of cities is highly restricted by the availability of data. Previous researchers looking at this study have mainly relied on statistical sources, like population census and economic data. A rigorous and sophisticated method using this kind of data for defining the polycentric structure was developed by McMillen (2001). He significantly advanced this field by adopting non-parametric techniques such as locally weighted regression (LWR) and semiparametric employment density functions to define subcenters as areas with significantly higher human densities than the expected density based on their distance from a CBD. This procedure has been widely used in subsequent work by McMillen (2003, 2004) and other researchers (Garcia-López, 2010; Riguelle et al., 2007). However, this method needs to subjectively select the CBD location, prior to the other processes. Identifying that location is very difficult for users who do not have

* Corresponding authors.
E-mail addresses: bohuang@cuhk.edu.hk (B. Huang), yimengsong@link.cuhk.edu.hk (Y. Song).

http://dx.doi.org/10.1016/j.rse.2017.06.039
0034-4257/

detailed knowledge of the study area; furthermore, main centers rarely have sharp region boundaries and the number of CBDs can change as a city grows. Those problems become more intractable when the study areas are huge and developing rapidly.

Besides, datasets like population censuses have a high level of accuracy and representativeness but a low update frequency, usually being renewed once every five or ten years. In addition, when such spatial statistics are used to determine the number of subcenters, the size of the observation units limits the adoption of the center definition method. For example, large units may produce fewer subcenter sites than more disaggregated data. Statistical data aggregated into administrative boundaries are unable to reveal the accurate distribution of human density below the administrative division level. When an administrative region is large, dense population sites may be ignored due to the large amounts of unused land within the same region.

Remote sensing data, like nighttime light satellite imagery, could also provide various features of urban landscape and infrastructure, adding a new potential source to study urban structure. For a long time, researchers have applied data from Defense Meteorological Satellite Program–Operational Linescan System (DMSP-OLS) to detect urban settlements (Elvidge et al., 1999, 2007; Ma et al., 2012; Sutton, 2003). Yu et al. (2014) developed an object-based method to characterize urban spatial patterns from nighttime light satellite images. As the new released data from Visible Infrared Imaging Radiometer Suite (VIIRS), more detailed inner-city structure monitoring became possible (Elvidge et al., 2013). Some studies have been carried out to estimate the socioeconomic indicators in finer resolution (Xi Li et al., 2013; Ou et al., 2015). The performance of urban area extraction at regional scale was verified by Shi et al. (2014). Although nighttime light data have relatively high spatial stability and guarantee the reliability of land parcel shaping, remote sensors are still incapable of recording socioeconomic attributes and human dynamics such as daily activities (Liu et al., 2015). For example, not only urban centers, but also roads, port regions, and industrial districts, emit glowing light at night, which can result in inaccurate estimates of population accumulation areas (Zhang et al., 2013).

In recent years, the rapid growth of location-based services and social media platforms has created new opportunities to discover the spatial characteristics of human behavior and activities (Jiang et al., 2016; Lee and Sumiya, 2010; Stefanidis et al., 2011). The high correlation between the check-in density of social media data and the human density distribution has been revealed by many studies (Cheng et al., 2011; Dunkel, 2015; Frias-Martinez et al., 2012; Steiger et al., 2015). Compared to conventional static data sources such as a population census, social media data are representative indicators with a much finer temporal-spatial scale that can depict the actual dynamics of the activities of urban dwellers (Hawelka et al., 2014). However, the check-in locations are so concentrated that most of the events occur around particular hot spots within a local region, resulting in serious spatial variability and regional instability. Therefore, there lacks a method that can provide both reliable land parcel shaping and quantitative analysis for the polycentric structure identification.

In this paper, we attempt to provide a method combining the advantages of nighttime light satellite images and social media check-in data for identifying the structure of polycentric cities. Three main steps are included: developing observation units, main center definition and subcenters definition. The Results section includes a test of our method across three cities and with different datasets and methods to demonstrate its effectivity. Two different methods are also provided to evaluate the accuracy of our results. The paper concludes with a summary of the method's advantages and the limitations of this study.

2. Study areas and data

2.1. Study areas

The three big cities, Beijing, Shanghai, and Chongqing, were selected as our study areas (Fig. 1). The geographical characteristics and urban morphology patterns vary greatly in these three cities, allowing us to verify the effectiveness and robustness of our method. Only municipal districts (city-controlled districts) are included in this study. Regions like the county-level cities or counties administratively belonging to these cities are removed.

As the capital of China, Beijing has undergone rapid suburbanization since the 1980s. It is densely populated with over 18.59 million residents (in 2015) living within the city-controlled districts of 12,046 km^2. The main built-up area of Beijing lies on the northern side of the North China Plain where the elevation ranges from 20 to 60 m. The plain topography means that the city's rapid expansion is not limited by natural conditions. The city has developed in a classic pie form, spreading out in concentric ring roads.

Shanghai is China's most populous city, where >24.15 million citizens (in 2015) are living in the city-controlled districts with area of 5462 km^2. The city sits in the Yangtze River Delta and the old urban and modern downtown are located on a vast alluvial plain that is divided by the Huangpu River. Due to the barrier of Yangtze River and the coastline, the city has mainly extended in the southwest direction.

Chongqing is the largest municipality under the direct administration of the Central Government and the only one in western China. The city-controlled area of Chongqing is about 15,162 km^2, with 15.7 population (in 2015) settling in. The city covers a large area crisscrossed by rivers and mountains, with great sloping areas at different heights. Therefore, the urban development is highly affected by topographical characteristics, leading to a relatively complex urban structure.

2.2. Data

2.2.1. Nighttime light imagery

In this study, nighttime light data are applied to characterize the textural features of urban built-up areas and to construct new observation statistical units. The Visible/Infrared Imager/Radiometer Suite (VIIRS), launched in October 2011, was designed to collect high-quality nighttime images in the day/night bands (DNBs), between 500 and 900 nm, with a ground spatial resolution of around 500 m (Miller et al., 2012). The VIIRS monthly composite data used in this study were obtained directly from the website of the Earth Observation Group, NOAA (http://ngdc.noaa.gov/eog/viirs/download_monthly.html), and the interference from stray light, lunar illumination, and cloud-cover is filtered out. The spatial resolution of the VIIRS product is 15 arc-seconds in the geographic grid covering the three cities' study areas.

2.2.2. Social media data

Compared to other spatial big data sources like mobile phone data, social media check-in record is believed to be more appropriate and commonly used for urban structure detection, because check-in events would normally be created when users are aware of something and stay in a particular position for relativity a long time (Kaplan and Haenlein, 2010). Weibo (microblogs) is one of the most popular social media platforms in China, whose monthly active users ("MAUs") reached 222 million in September 2015 and mobile MAUs represent 85% of the total MAUs (Weibo Corporation, 2015). In this study, the check-in map of Weibo is used as representative spatial information on human activities.

3. Methodology

3.1. Data preprocessing

Thirteen images covering the three study areas from April 2015 to April 2016 were collected for this study. An averaging map was calculated from those images to reduce noise and to use in null pixels. Finally, the data were projected into the Universal Transverse Mercator (UTM) grid with a resolution of 500 m.

With the movement of crowds and dynamic hotspots of interest, the check-in distribution of Weibo changes over time. The spatial

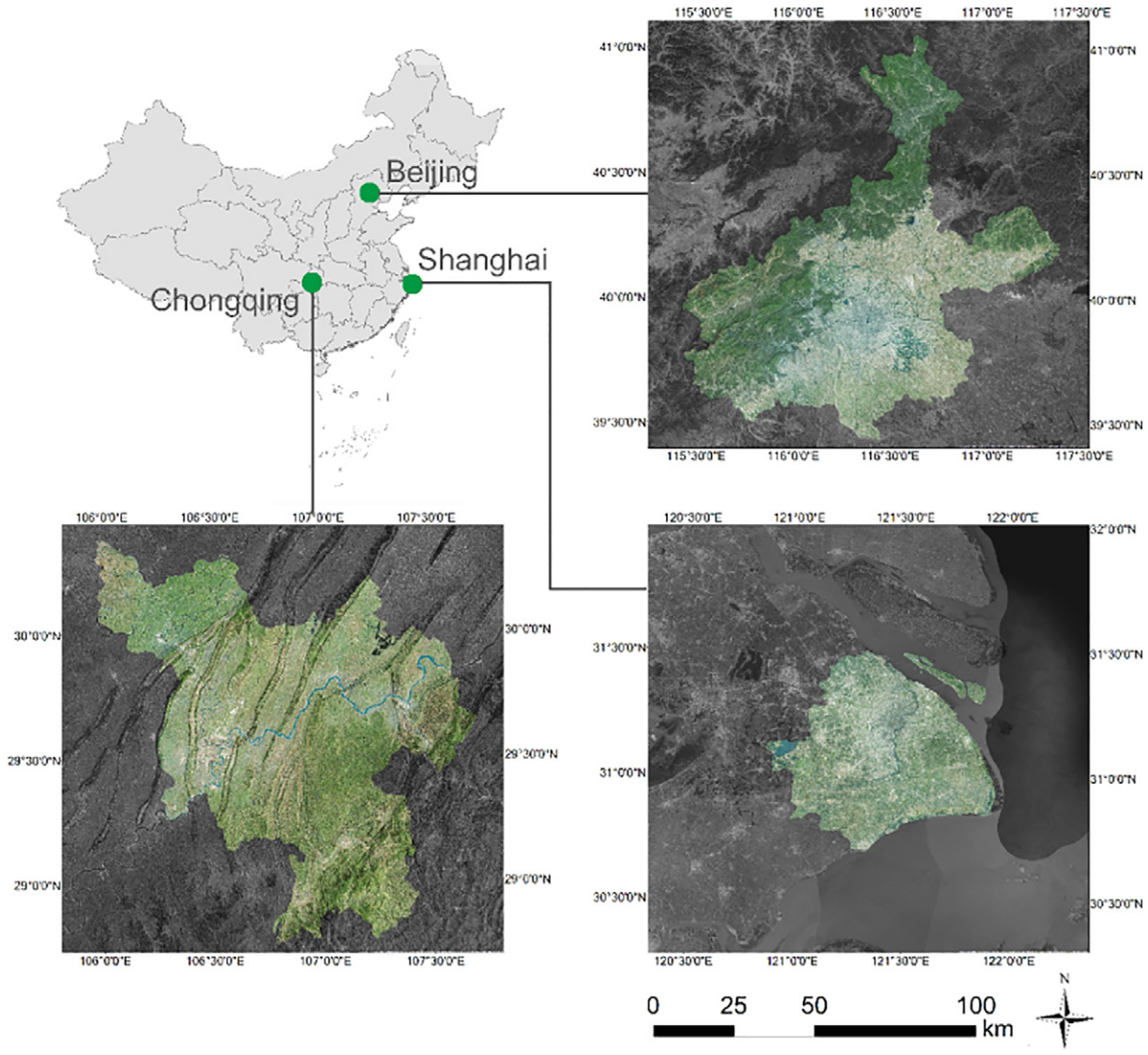

Fig. 1. Spatial organization of the three study areas. The colorized regions are the three cities respectively: Beijing, Shanghai and Chongqing.

stability of this data along a temporal dimension is an important issue when it is used to uncover the urban structure, which is regarded as constant during a short period. To overcome this problem, this study uses a dataset of >5.6 million check-in observations through the Weibo Application Programing Interface (API) (http://open.weibo.com/wiki/API%E6%96%87%E6%A1%A3_V2/en) from April 25, 2015 to May 25, 2016 (396 days), and the amount of check-in records was calculated for each grid using the same UTM projection as for the nighttime light data.

3.2. Developing new observation units

Active human movements and business activities may be concentrated in a small part of an administrative district, and may easily spread from one district to a neighboring district; they are not limited by administrative boundaries. To determine the spatial characteristics of social activities, new observation statistical units are essential. Here, the object-oriented segmentation methodology is applied to create observation units which can capture the distribution of human activities. Several segmentation approaches have been proposed in the past decades, such as mean-shift segmentation (Comaniciu and Meer, 2002), and the fractal net evolution approach (Baatz et al., 2006).

In this study, the image segmentation is performed on nighttime light data and under the multiagent object-based classification framework (MAOCF) (Zhong et al., 2014), which could optimally control the procedure of object merging and utilize the contextual information from the surrounding objects. Three key parameters: scale factor, shape factor and compactness factor, are set to control the segmentation process. Among these parameters, the shape, and compactness factors are relatively easy to determine; the former is set to 0.2 to emphasize the spectral importance, and the latter to 0.7 to maintain the created objects more aggregated and lower perimeter/area ratio. The scale parameter represents the maximum allowable heterogeneity of an image object and determines the object's size and should match the level of spatial detail required for a specific study. Thus, an appropriate scale factor method based on segmentation's quality criteria can be better than data from an empirical setting. Although nighttime light data have higher spatial stability and can guarantee reliable land parcel shaping, which work well for segmentation, the “blooming” effect (Li and Zhou, 2017) makes this type of data less effective for presenting an inner-city structure. In addition, spaces that emit bright light are not always areas with a high density of human activity, because areas such as manufacturing districts and cargo terminals can emit bright light without attracting much population flow (Zhang et al., 2013), whereas the

check-in records collected by social media can clearly reveal the spatio-temporal distribution of human activities (Jiang et al., 2012). Consequently, evaluation of the segmentation scale is based on social media data.

The main goal of segmentation is to maximize intra-segment homogeneity and maximize inter-segment heterogeneity of the check-in density. An optimization method (Chabrier et al., 2006) based on segmentation quality criteria, was applied in this study. The intra-segment homogeneity can be calculated using the global weighted variance (wVar), whereas the inter-segment heterogeneity can be calculated using the spatial autocorrelation measure, Global Moran's I (GMI) (Espindola et al., 2006).

$$wVar = \frac{\sum_{i=1}^{n} a_i v_i}{\sum_{i=1}^{n} a_i}, \tag{1}$$

where v_i is the variance of check-in density in segment i, and a_i represents the area of the segment.

$$GMI = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i-\bar{x})(x_j-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2\left(\sum\sum_{i\neq j} w_{ij}\right)}, \tag{2}$$

where n is the total number of segments, x_i and x_j are the mean check-in density of segments i and j, respectively, and $\bar{x}$ is the mean check-in density of the whole area. The weighting value w_{ij} is measured based on the inverse distance between the geometric center of segments i and j.

Given the effect of intra-segment homogeneity and inter-segment heterogeneity, a combination goal (CG) is applied (Zhou et al., 2014). The optimized scale factor is achieved when CG reaches the minimum.

$$CG = V_{norm} + MI_{Norm}, \tag{3}$$

where V_{norm} and MI_{norm} are the normalized $wVar$ and GMI, respectively.

To maintain spatial stability and restrict the random error of check-in location, all of the segmentation units not >1 km^2 are merged with the smallest neighboring units.

3.3. Main center definition

A main center can be defined as a large area with high population density that has the characteristics of a spatial cluster. Basically, main center is located at the core of a city and covers the central business district (CBD). Therefore, the Anselin Local Moran's I (Anselin, 1995) is applied to find the main center of each city. The Local Moran's I (LMI) for the ith segmentation unit is given as

$$LMI_i = \frac{x_i-\bar{x}}{S_i^2}\sum_{j=1,j\neq i}^{n} w_{ij}(x_j-\bar{x}), \tag{4}$$

where all of the variables have the same meaning as in Eq. (3), and S_i^2 is the global sample variance such that

$$S_i^2 = \frac{\sum_{j=1,j\neq i}^{n}(x_j-\bar{x})}{n-1}, \tag{5}$$

A positive value for LMI_i indicates that i has neighbors with similar values; therefore, segment i and its part neighbors can form a cluster (Anselin, 1995). To pick up all of the segments with statistically significant positive LMI values, a z-score is introduced (Mitchell, 2005):

$$z_{LMI_i} = \frac{LMI_i - E[LMI_i]}{\sqrt{V[LMI_i]}}, \tag{6}$$

where

$$E[LMI_i] = -\frac{\sum_{j\neq i}^{n} w_{ij}}{n-1}, \text{and} \tag{7}$$

$$V[LMI_i] = E\left[LMI_i^2\right] - E[LMI_i]^2. \tag{8}$$

A high positive z-score (larger than 1.96) for a segment indicates that it is a statistically significant (0.05 level) spatial outlier. Those segments with high values that are surrounded by other segments with high values (HH) are defined as main centers, due to the distinctly crowded check-in density over the whole area. When more than one clusters of segments are included in the candidate parts of the main center, the largest one with more check-in records will be identified as the main center area.

Next, we define the check-in density weighted centroid of the main center as the center point of the city. This city-center point will be used to calculate the distance between each segment and the main center of the city.

3.4. Subcenter definition

A subcenter is defined as a set of contiguous tracts with high levels of human activities density. There are other components of the intra-city polycentric structure, which include edge cities and satellite towns. Qualified subcenters should have significantly higher densities than their immediate surroundings (locally high), and relatively high human activity density compared to all of the segments in the study area (globally high). A two-step strategy is applied to identify the locations of subcenters.

First, we define the check-in density weighted centroid of the main center as the central point of the city (city-center point). Then, a geographically weighted regression (GWR) procedure is used to model the relationship between the distance to the city-center point from the geometric center of an individual segmented region and the square root of its check-in density value. Theoretically, the closer to the city-center point, the more human activities there will be. However, this pattern cannot be examined with an ordinal logistic regression (OLR), due to the asymmetric distribution of the subcenters (McMillen, 2001). The development of subcenters is determined by the social and natural geographical characteristics of a city with strong spatial nonstationary. GWR provides a local modeling tool to fit a regression equation to every observation in the dataset. As only nearby observations are used in such estimations (Fotheringham et al., 2003), local rises in check-in density are expressed as positive residual errors. The GWR formula is given as

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)d_{ik} + \varepsilon_i, \tag{9}$$

where y_i is the estimated square root of the check-in density for segment i; u_i and v_i denote the spatial center; β_0 (u_i, v_i) is the intercept, β_k (u_i, v_i) is the local estimated coefficient of the kth independent variable for the segmentation unit i; and ε_i is the residual error. The Gaussian kernel is constructed as an adaptive distance, and the cross validation (CV) method is used to select an optimal bandwidth. The subcenter candidates are those sites with standard residuals >1.96, implying that their check-in density values are significantly higher than average at the local scale.

Next, after the subcenter candidates are selected with the significant positive residual errors, an overall classification is applied to remove candidates whose actual human activity or area is lower than that of the other units in the whole study area. Jenks's natural breaks classification (NBC) (Jenks, 1967) is a data-clustering method that can be used to reduce the variance within classes and maximize the variance between classes. By using NBC, we divided all the tracts of each study area into

several classes according to check-in density and area, based on the goodness of variance fit (GVF) value. A GVF value >0.8 is the generally accepted threshold to evaluate the classification result (Jenks, 1967). The number of classes grew from two and stopped when the GVF exceeded 0.8. Tracts that belonged to the class with the lowest level of check-in density or the smallest area were not considered to be subcenters. Finally, qualifying subcenters were neither within the main center, as defined before, nor adjoining any main center regions.

4. Results

4.1. Segmentation

There are 866, 927, and 558 units generated by segmentation in the three study areas, as shown in Fig. 2. Significant variation in the segment sizes between urban and exurban areas is common in Beijing, according to the enlarged views. The denser areas are described in more detail with finer segmentation, whereas in exurban zones, the units are relatively coarser. This phenomenon may be the result of changes in the spatial characteristics of nighttime lights and the social media check-in density, which are more frequent in the plains but rather rare in mountainous regions; thus, to satisfy the variance accumulation needed to form a unit in the segmentation process, a larger area is required in regions with nighttime light intensity. Similarly, in the eastern part of Chongqing, the mountainous areas with fewer human activities result in larger land parcels than the variegated parcels on the plains. However, in the most populous city, Shanghai, such differentiation is not so obvious, and human activities are more evenly distributed throughout the city, which occupies a great flood plain. The parcel scale of the urban areas in Shanghai is as fine as those in Beijing and Chongqing, but in the exurban areas, the approximate value is less than double that of the urban areas. Due to its location on an alluvial plain, human activities in Shanghai are more evenly distributed, resulting in more uniform segment sizes.

Although the total municipal district area of Chongqing (15,162.16 km^2) is greater than those of Beijing (12,046.17 km^2) and Shanghai (5462.85 km^2), there are fewer land parcels in Chongqing than in the other two cities (558 compared to 866 and 927, respectively) (Table 1). The segmentation patterns show that these three cities have diverse urban structures and population distributions. Very small units (smaller than 1 km^2) are merged to the nearest one.

4.2. Main center

The main centers of the three cities are explored using the Local Moran's I value. As discussed in the Methodology section, sites with high values that are surrounded by other segments with high values (HH) are selected as candidates for the main center. All of the

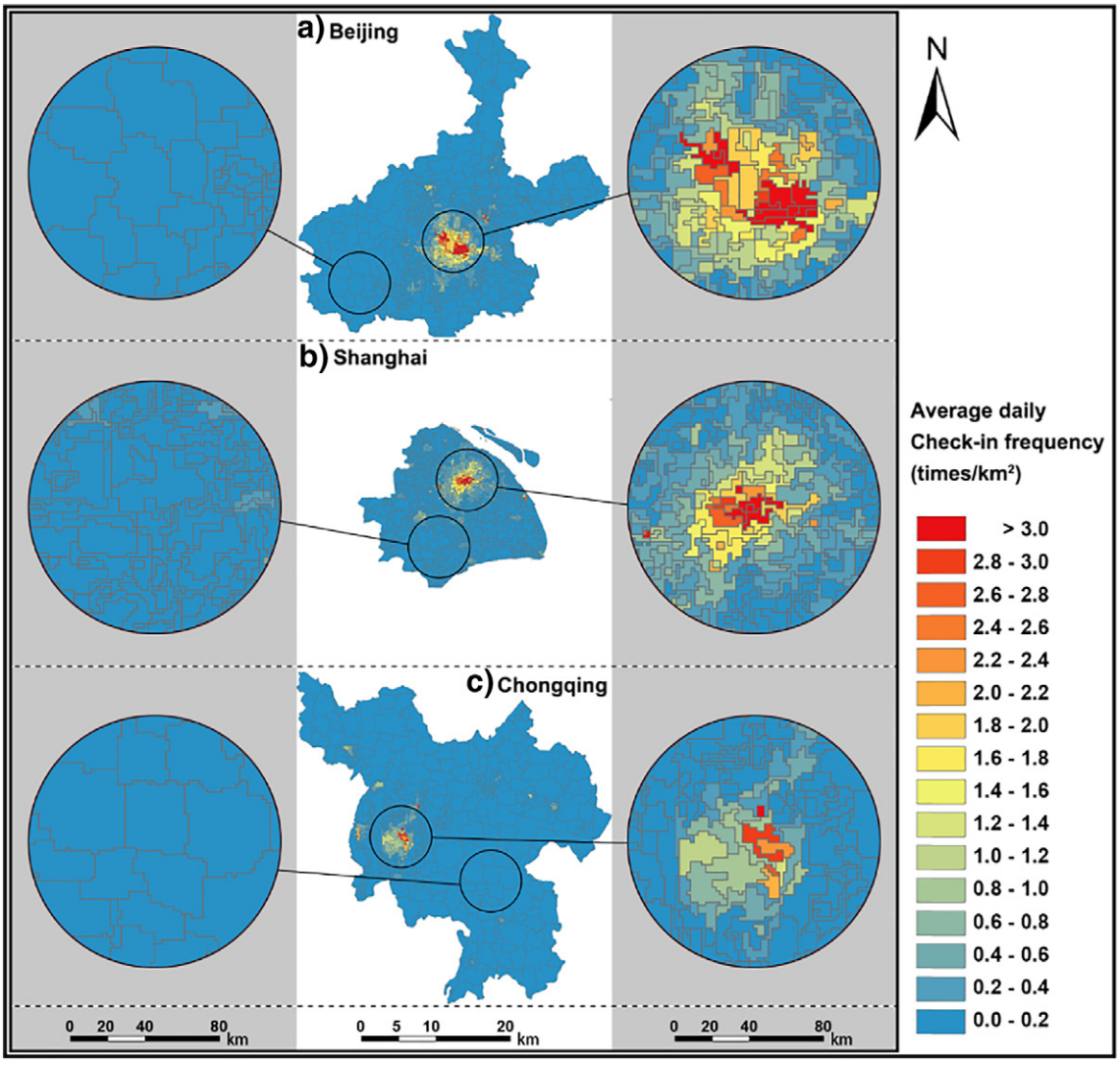

Fig. 2. Check-in density statistics in newly formed observed units. The enlarged views of urban and exurban areas are given to the right and left of the main maps, respectively.

geographically contiguous candidate parts are merged together, and the spatial outlier candidate sites are excluded. As Fig. 3 shows, the main center of Beijing is located in the middle of the plain and has an area of 688.5 km^2; it covers 5.72% of the study area and includes 73.78% of the check-in records. The main center of Chongqing is 404.8 km^2, and is located in the west of the city; it covers 2.70% of the study area and includes 67.44% of the check-in records. The main center of Shanghai is mid-sized, covering 449.0 km^2; it is located in the northeast of the city covering 8.22% of the urban area and accounts for 60.76% of the check-in records.

4.3. Subcenter

The first panel of Table 2 presents the results of the geographically weighted regression (GWR) estimates, with the dependent variable being the square root of the check-in density and the explanatory variable being the distance from the center point. The GWR estimates 64 significant positive residuals in Beijing, 48 in Shanghai, and 38 in Chongqing. The geographically contiguous land parcels are all merged together, and all land parcels that belong to the last class of the NBC are excluded. To separate the subcenters from the main centers, the land parcels within or contiguous with the main center areas or with the lowest check-in density were deleted. Candidate subcenters with lowest area (NBC) compared to all land parcels were also excluded. We finally obtained 10 subcenters in Beijing, 12 in Shanghai, and 8 in Chongqing. The final subcenters and their locations in the three cities are displayed in Fig. 3.

5. Comparison and evaluation

To demonstrate the effectiveness of the proposed method, comparative experiments were conducted with two types of data and three methods. The urban polycentric structure was also assessed by the spatial delineation accuracy and detection accuracy (Taubenböck et al., 2013).

5.1. Comparison experiment

Nighttime light data have been widely applied to discover the urban structure, but tracking human activities is difficult. To better demonstrate the validity and superiority of the combination with social media data, comparisons were conducted by taking the mean value of nighttime light intensity for each segmentation unit as the input for center detection instead of social media data.

The relative cut-off threshold method was commonly applied in previous studies because of its simple operation and greater objectivity compared to the use of an absolute threshold (Small et al., 2005). This method was applied as one of the comparative methods to detect the urban structure. As the first comparative method, followed by Liu and Wang (2016), the 90th-percentile of the highest nighttime light intensity or check-in density segment units by area in China's megacity were defined as urban centers (selecting the top 10% areas with the highest nighttime light intensity or greatest density of check-in units). The second comparative method was suggested by McMillen (2003), who applied the ordinary least squares (OLS) method and indicated that the population density declined as the distance to the subcenters increased. Because the OLS method is unable to define the main center for polycentric cities, the same processing of the main center definition by LMI clustering of our method was applied in this experiment. Thus, six sets of comparative experiments, including our method, and data were applied to the three study areas.

Fig. 4 shows the results of the comparisons. Using a cut-off threshold method, the main center and subcenter cannot be distinguished because of their unclear boundaries. The results with the threshold are shown in orange, and for the other two methods, the main centers are shown in red and the subcenters in yellow. As a coastal city, the nighttime light of Shanghai is greatly disturbed by ports and industrial areas, so the threshold method cannot distinguish the complete main center part from the nighttime light data (b1). Similarly, it should be noted that by using nighttime light data only there is no dominant continuous main center region in Shanghai under LMI clustering (b3, b5), which indicates the poor performance of nighttime light in this case. A clear main center boundary could hardly be detected by the threshold method, especially when only nighttime light data were applied. In Beijing, the main center area tends to spread into the suburban area, and there are too many fragments in a1 and jagged shapes in a2. From the results of Shanghai, the main center and subcenters are the easiest to distinguish from the threshold method on social media data (b2), but estimation from nighttime light has poor results, and the area of the main center cannot be estimated (b1). We defined the top 10% of areas with highest nighttime light intensity or check-in density as the center areas in this comparative study, but this value does not fit well across different regions. In Chongqing, too many regions are classified as center parts, even remote regions with high coverage of natural lands (c1, c2).

LMI works better than the threshold method in detecting the main center, except for the use of nighttime light data in Shanghai (b3, b5). The outlines of the main centers in Beijing and Chongqing according to the results of nighttime light and Weibo check-in data are similar. However, the subcenter detection results vary considerably. With OLS, there are more outliers with significant positive residuals in the center and covered by the main center or in very remote places (a4) because the global model cannot perform well in local regions. By comparing the results of Shanghai, more regions along the edge of the administrative boundary were detected as subcenters by OLS, whether based on nighttime light or Weibo data (b3, b4). Only the result in b6 can better detect both the complete main center and the subcenters, which are evenly distributed in the plain area of Shanghai. In Chongqing, similar patterns were detected with OLS and GWR from Weibo data (c4, c6), whereas fewer subcenters can be identified with nighttime light data (c3, c5).

5.2. Delineation accuracy

We introduced a quantitative analysis to verify the spatial coverage accuracy of the proposed data processing and method compared to other datasets and approaches.

Determining the accurate distribution of an urban center map is difficult. We evaluated the agreement between the coverage of the detected center regions and the coverage of points of interest (POI) data. POIs represent locations with a series of information, including land use category, geographic location, and other geographic features, which is useful to help identify the scope of functional urban area (Hu et al., 2016). Compared to the urban impervious surface data, POI data could better represent an area with greater human activity and exclude industrial regions.

We accessed 17 types of POI that are highly related to human daily life from Google Maps API (https://developers.google.com/places/web-service/intro). The numbers of each type of POI in the three study areas are shown in Table 3. The POI data were plotted by the coordinates onto a gridded map with the UTM projection and 500-m resolution, as

Table 1
Segmentation statistics.

		Beijing	Shanghai	Chongqing
Land parcel number		866	927	558
Area (km^2)	Total	12,046.17	5462.85	15,162.16
	Max	221.39	56.75	200
	Min	1.00	1.00	1.00
	Mean	13.91	5.89	27.17
	SD	22.84	7.03	34.23

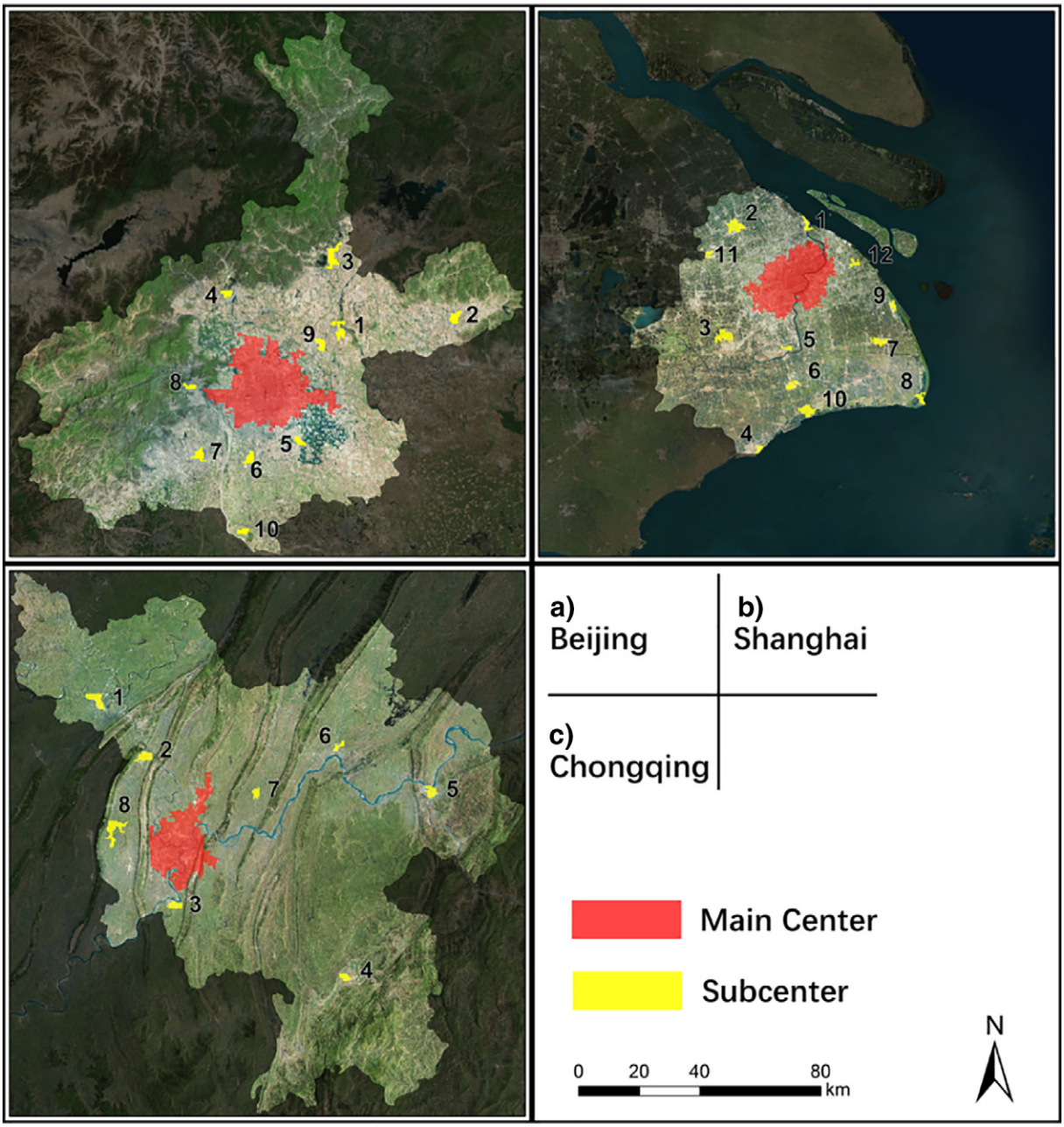

Fig. 3. Main center and subcenter identification with satellite base map.

with the other prepared data. POI coverage bitmaps of different types were generated by determining whether each grid contained one or more POI records. Finally, the POI coverage map was applied to compare the results for the center area. We used Cohen's kappa to measure the distribution agreement between the POI coverage maps and the center results. Although no single type of POI can represent the true distribution of the urban center, the greater agreement between the daily life–related POI coverage and the center-detected map indicates greater accuracy of the results.

The results of agreement presented by kappa coefficient in the three study areas are listed in Tables 4–6. The results of social media check-in data and remote sensing data under the processing of LMI clustering and GWR perform better than those in other cases in matching the coverage of the POIs.

Table 2
Estimation results for subcenters.

		Beijing	Shanghai	Chongqing
GWR estimation	Number of significant GWR residuals	64	48	38
	Number of land parcels within or contiguous with the main center	24	13	15
	Number of land parcels belonging to the last class of NBC* (either density or area)	13	9	4
Subcenter	Number of land parcels belonging to subcenters	27	24	19
	Number of subcenters	10	12	8

NBC*: natural breaks classification.

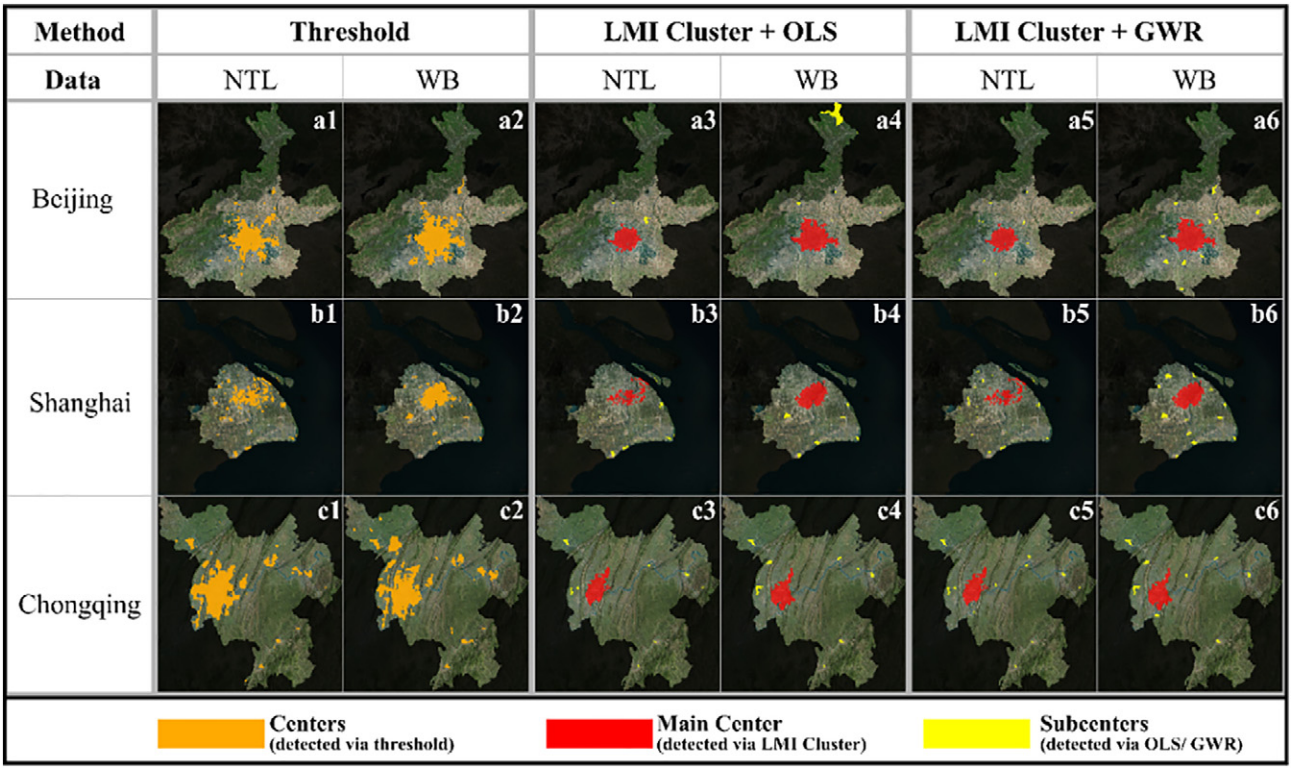

Fig. 4. Centers detected by different methods and datasets. All center areas detected by threshold are shown in orange. In the results from the other two methods, the main centers are shown in red and the subcenters in yellow. "NTL" indicates that nighttime light data are involved in detecting centers, whereas "WB" indicates that the Weibo check-in data are used to detect centers.

5.3. Detection accuracy

The detection accuracy evaluates the performance of the developed method to detect the urban polycentric structure. For this purpose, we collected the three cities' master plans (the only legal documents) for 2020 and compared our results with the centers defined by the governments at the district level. Overall, our method has good performance in identifying of both the main center and the subcenters (Table 7). More details of the detection accuracy evaluation are given below.

a) Beijing: The master plan of Beijing 2004–2020 (Beijing Government, 2005) defined the main center as six districts (Fig. 3a), all of which were detected with our method. Two of the 11 subcenters in the master plan (Miyun and Yanqing) are not within our study area; the other nine (Fig. 3a, No. 1–9) were all detected (Tongzhou was included in the main center in our result because of urban expansion of the main center). In addition, our result also includes another two subcenters: the airport with its surrounding area and Huanggezhuang. Therefore, the user's accuracy is 88.2%.
b) Shanghai: Eleven subcenters are listed in the recent master plan of Shanghai for 2020 (Shanghai Government, 2001). One subcenter (Chengqiao) is not within our study area; nine of the other 10 were identified (Fig. 3b, No. 1–9). The Qingpu subcenter was not included in our results. The master plan defined six districts as the

Table 3
The number of POI accessed for three study regions.

POI type	Beijing	Shanghai	Chongqing
ATM	5068	4747	2425
Bakery	1145	1303	541
Bank	3288	3485	1738
Beauty salon	725	725	414
Clothing store	1540	1983	915
Convenience store	8036	9632	2224
Haircare	2489	1981	820
Home goods store	1514	1816	798
Hospital	1932	1502	1238
Laundry	1100	1104	495
Local government office	2453	1860	1013
Lodging	2405	1984	1246
Pharmacy	2078	1783	1757
Police station	1524	1767	991
Post office	861	743	773
School	2078	1812	1421
Shopping Mall	682	665	412

Table 4
Kappa coefficient represents the agreement of detected center area with the coverage of different types of POI (Beijing). The numbers in bold represent the best results for each POI type.

POI type	Threshold		LMI cluster + OLS		LMI cluster + GWR	
	NTL	WB	NTL	WB	NTL	WB
ATM	0.44	0.439	0.425	0.332	0.427	**0.486**
Bakery	0.178	0.179	0.179	0.122	0.181	**0.206**
Bank	0.351	0.342	0.342	0.25	0.342	**0.391**
Beauty salon	0.109	0.11	0.111	0.073	0.115	**0.131**
Clothing store	0.147	0.147	0.147	0.1	0.152	**0.174**
Convenience store	0.354	**0.397**	0.258	0.249	0.263	0.368
Haircare	0.187	0.198	0.168	0.114	0.169	**0.205**
Home goods store	0.13	0.131	0.094	0.075	0.1	**0.142**
Hospital	0.118	0.124	0.118	0.076	0.12	**0.142**
Laundry	0.174	0.18	0.175	0.114	0.179	**0.2**
Local government office	0.108	0.102	0.102	0.06	0.108	**0.118**
Lodging	0.165	0.166	0.162	0.106	0.164	**0.178**
Pharmacy	0.163	0.168	0.159	0.11	0.162	**0.191**
Police station	0.172	0.19	0.131	0.112	0.134	**0.195**
Post office	0.118	0.115	0.128	0.078	0.127	**0.142**
School	0.152	0.163	0.133	0.093	0.135	**0.173**
Shopping mall	0.092	0.093	0.087	0.058	0.089	**0.112**

Table 5
Kappa coefficient represents the agreement of detected center area with the coverage of different types of POI (Shanghai). The numbers in bold represent the best results for each POI type.

POI type	Threshold		LMI cluster + OLS		LMI cluster + GWR	
	NTL	WB	NTL	WB	NTL	WB
ATM	0.225	0.394	0.205	0.213	0.388	**0.397**
Bakery	0.087	0.177	0.081	0.089	0.169	**0.184**
Bank	0.156	0.281	0.153	0.157	0.277	**0.288**
Beauty salon	0.053	0.092	0.052	0.064	0.082	**0.103**
Clothing store	0.072	0.123	0.071	0.079	0.110	**0.125**
Convenience store	0.180	0.317	0.134	0.142	0.302	**0.323**
Haircare	0.094	0.172	0.083	0.094	0.150	**0.176**
Home goods store	0.046	0.110	0.031	0.038	0.089	**0.112**
Hospital	0.052	0.101	0.039	0.044	0.087	**0.108**
Laundry	0.091	0.180	0.085	0.091	0.173	**0.183**
Local government office	0.078	0.125	0.071	0.077	0.119	**0.136**
Lodging	0.105	0.161	0.080	0.083	0.140	**0.163**
Pharmacy	0.077	0.160	0.066	0.071	0.150	**0.171**
Police station	0.048	**0.092**	0.034	0.040	0.074	0.091
Post office	0.065	**0.113**	0.076	0.080	0.109	0.111
School	0.091	0.164	0.075	0.083	0.159	**0.171**
Shopping mall	0.028	0.065	0.025	0.032	0.058	**0.072**

main center, which is same as our results (Fig. 3b). Moreover, three other subcenters (Fig. 3b, No. 10–12) were also detected with our method. This represents a user accuracy of 83.3%.

c) Chongqing: The master plan of Chongqing 2007–2020 (Chongqing Government, 2007) pointed out nine districts as its main center; eight of them were defined as the main center (Fig. 3c) in our results, and one was defined as a subcenter (Fig. 3c, No. 2). The five subcenters (Fig. 3c, No.1–2, 4–6) listed in the master plan were all found with our method. Two regions that belong to the main center were defined as subcenters because of their distance from the main center. The results of Chongqing give a user accuracy of 86.7%.

6. Discussion

6.1. Spatial stability of check-in data

Randomness exists in human behavior; therefore, the check-in distribution of Weibo should vary with time and affect the stability of the input data. To overcome this randomness, we accumulated the social media data up to 396 days, with 5.6 million check-in records in these three cities, to enhance the data's spatial stability. Over time, the proportion of any region's check-in record number to the whole study area's check-in record number would stabilize. Thus, clear and reliable variance across the whole study region can be more easily detected with the spatial statistics model. To measure how the spatial stability would change over time, we introduced an index:

$$S^t = \frac{\sum_{i=1}^{j}\left|\frac{\sum_{k=1}^{t} a_{ik}}{\sum_{k=1}^{t} Q_k} - \frac{\sum_{k=1}^{t-1} a_{ik}}{\sum_{k=1}^{t-1} Q_k}\right|}{j}, (t \geq 2), \quad (10)$$

where a_{ik} is the check-in records of grid i and day k, $\sum_{k=1}^{t} a_{ik}$ represents the cumulative check-in records from the 1st day to the tth day within grid i, $\sum_{k=1}^{t} Q_k$ represents the cumulative check-in records from the 1st day to the tth day within the entire study area, j is the total number of grids that cover study area, $\frac{\sum_{k=1}^{t} a_{ik}}{\sum_{k=1}^{t} Q_k}$ is the ratio of total check-in records from the 1st day to the tth day in one grid to that in the entire study area, Theoretically, for any grid k, $\frac{\sum_{k=1}^{t} a_{ik}}{\sum_{k=1}^{t} Q_k}$ would become stable as the longer of cumulative days (the larger of t). S^t is designed to measure the averaging absolute of the difference between t days and $t-1$ days of accumulation for the whole image. The smaller the S^t, the higher the spatial stability.

As Fig. 5 shows the three curves of Beijing, Shanghai, and Chongqing level off or decrease slowly after 100 days, which suggests that after a relatively long time, the daily average check-in records in each grid will stabilize, and 396 day's data is more than enough for this study. Besides, by comparing the three cities, the three cities have similar fluctuation trend of the social media records, which may be affected by social events and holidays. Among those, Shanghai, with the most populous citizens, has the highest value of the overall variance of check-in records, while Chongqing is the least.

Although it has little effect, bias still exists because social media data are regarded as non-representative data (Zagheni and Weber, 2015). Weibo check-in records tend to leave out some sections of society because children, the elderly, and the poor are less-frequent Weibo users. Nevertheless, such data can still represent the population distribution pattern within cities; the correlation between social media data and population distribution has been verified with survey data (Liu and Wang, 2015). As more studies of the representativeness are carried out, the reliability and accuracy of this method will improve.

Table 6
Kappa coefficient represents the agreement of detected center area with the coverage of different types of POI (Chongqing). The numbers in bold represent the best results for each POI type.

POI type	Threshold		LMI cluster + OLS		LMI cluster + GWR	
	NTL	WB	NTL	WB	NTL	WB
ATM	0.440	0.439	0.425	0.427	0.332	**0.486**
Bakery	0.280	0.302	0.320	0.316	0.355	**0.357**
Bank	0.121	0.134	0.142	0.138	0.159	**0.164**
Beauty salon	0.205	0.223	0.239	0.236	0.264	**0.265**
Clothing store	0.081	0.089	0.098	0.097	0.115	**0.118**
Convenience store	0.102	0.104	0.121	0.119	0.125	**0.129**
Haircare	0.277	0.317	0.301	0.297	0.362	**0.363**
Home goods store	0.146	0.157	0.152	0.153	0.179	**0.187**
Hospital	0.115	0.128	0.133	0.131	0.159	**0.163**
Laundry	0.124	0.140	0.126	0.125	0.149	**0.152**
Local government office	0.126	0.147	0.143	0.142	0.175	**0.178**
Lodging	0.108	0.119	0.112	0.110	0.130	**0.134**
Pharmacy	0.093	0.101	0.122	0.119	0.140	**0.141**
Police station	0.142	0.150	0.139	0.137	0.160	**0.164**
Post office	0.154	0.170	0.156	0.153	0.193	**0.198**
School	0.082	0.094	0.092	0.090	0.110	**0.114**
Shopping mall	0.139	0.151	0.123	0.122	0.152	**0.157**

6.2. Segmentation

Segmentation is vital to form the new observation units because the administrative boundary is too coarse for the detection of human activities as discussed above. In addition, compared to raster-based processing, processing based on object-oriented segment units from nighttime light data could easily observe the clear boundary of the center so that fewer fragments would be created. In this study, to be noticed, a proper size of segmentation should be well considered, which is controlled by the scale factor. In this study, a multiresolution operated multiagent object-based classification framework (MAOCF) has the advantage of increased heterogeneity within individual segmentation units combined with the compactness of units over the whole study area. The combination goal (CG) is introduced that controls the selection of the optimized scale factor by its two components: global weighted variance and Global Moran's I. As the scale factor grows, the variance within each segment would generally increase because more cells are included in one unit, whereas the mean values from all segment units become less diverse, which causes global Moran's I to decrease. Such processing is self-adaptive in that the optimized scale factors for the three study cities range from 5 to 7, as shown in Fig. 6.

Table 7
Detection accuracy of Beijing, Shanghai, and Chongqing.

		Master plan	Our results	User's accuracy
Beijing	Main center	Dongcheng, Xicheng, Chaoyang, Fengtai, Shijingshan, Haidian	Dongcheng, Xicheng, Chaoyang, Fengtai, Shijingshan, Haidian, Tongzhou[b]	88.2%
	Subcenters	Shunyi, Pingguo, Huairou, Changping, Yizhuang, Daxing, Fangshan, Mentougou, Tongzhou	Shunyi, Pingguo, Huairou, Changping, Yizhuang, Daxing, Fangshan, Mentougou, Airport[a], Huanggezhuang[a]	
Shanghai	Main center	Huangpu, Xuhui, Changning, Jingan, Putuo, Hongkou, Yangpu	Huangpu, Xuhui, Changning, Jingan, Putuo, Hongkou, Yangpu	83.3%
	Subcenters	Baoshan, Jiading, Songjiang, Jinshan, Minhang, Nanqiao, Huinan, Nanhui, Airport	Baoshan, Jiading, Songjiang, Jinshan, Minhang, Nanqiao, Huinan, Nanhui, Airport, Haiwan[a], Anting[a], Caolu[a]	
Chongqing	Main center	Yuzhong, Dadukou, Jiangbei, Shapingba, Jiulongpo, Nanan, Yubei, Banan, Beibei	Yuzhong, Dadukou, Jiangbei, Shapingba, Jiulongpo, Nanan, Yubei, Banan (partly)	86.7%
	Subcenters	Hechuan, Nanchuan, Fuling, Changshou, Longxing	Hechuan, Beibei, Banan (partly)[a], Nanchuan, Fuling, Changshou, Longxing, Huxi-Chenjiaqiao[a]	

[a] Centers defined by the government and by our method do not match.
[b] Subcenter covered by main center.

In our experiments, it was found that the final detection results would not change much if the optimal value of scale factor was adjusted by adding or subtracting by 1, because the segmentation algorithm would generate smaller units in urban central areas, and the units under 1 km^2 would be merged to their nearest one. However, very small scale factor would lead to many fragments in the final urban centers so that clear center boundaries cannot be observed, whereas very big scale factor would merge land parcels into larger ones so that high check-in density spots would become less significant.

6.3. POI as validation source

Unlike land cover types, the center or subcenter region is a concept that describes a concentrated place with many human activities, which differs from an impermeable surface. To verify the results in this study, we accessed a series of POIs from the Internet. The POI data have been proved effective in detecting different land use types (Hu et al., 2016). However, no single type of POI can represent the location of urban centers because of the distribution inequality or information collection bias. More importantly, POIs are not updated in a timely manner, and users can hardly assess the generating time of the POI data or confirm whether a POI facility remains valid. In this study, we applied up to 17 types of POI collected very recently and calculated the agreement of the POI coverage with the urban center and subcenter results; the weakness of the POIs can be reduced because the horizontal comparison was conducted.

7. Conclusions

In this study, the polycentric structures of three cities were identified using new types of data, including social media check-in maps and nighttime light images with a high spatial resolution. A series of quantification methods were used, and the results of the different cities were compared. The results of the comparative experiment with different methods and two different types of accuracy evaluation demonstrate the robustness of the proposed method. This method could be easily extended to the urban structures of other rapidly developing cities, even if the users have inadequate background knowledge of the cities they are studying.

In the three cases examined in this study, the main centers and the subcenters with large amounts of urban human activity were identified with clear boundaries. Unlike previous research methods based on census data, our approach overcomes the limitations of statistical data organized by administrative units and describes the areas of human activity with new statistical units that produce more accurate results. We first applied segmentation of nighttime light images to reform the statistical units in the urban area. In this era of big data, social human activity data, such as social media data, mobile phone records, and taxi trajectories, can be used to measure the actual dynamics of urban populations at a much finer spatiotemporal scale than is possible with conventional static data sources. Moreover, the comprehensive use of various types of data creates more opportunities for future research.

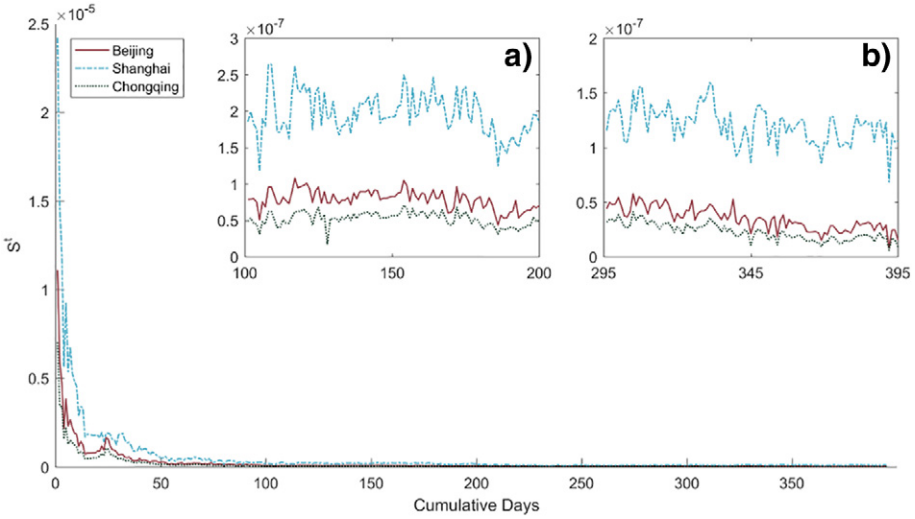

Fig. 5. Spatial stability changed over time. Subplots a and b are the enlarged views of days 100 to 200 and the last 100 days, respectively.

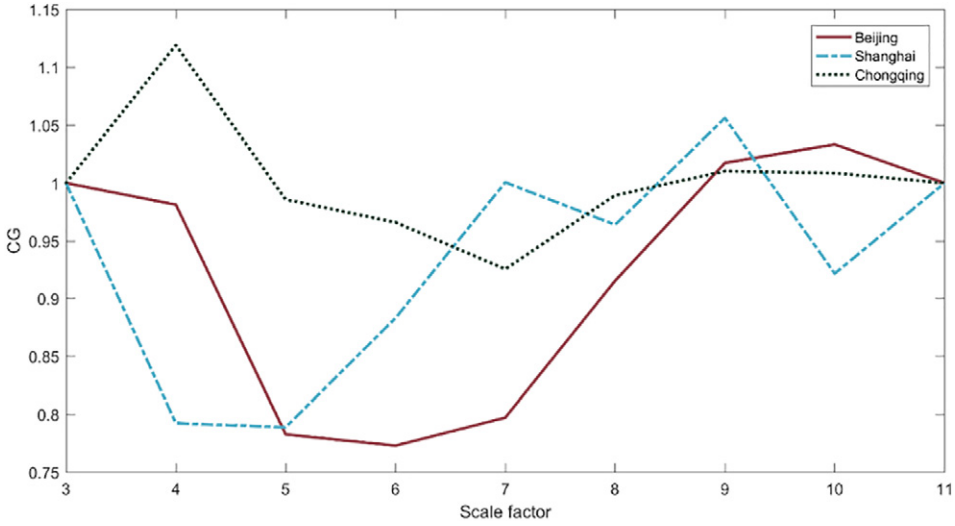

Fig. 6. Dynamic combination goal with different selection of scale factor.

Acknowledgments

This work was funded by the Hong Kong Research Grants Council under GRF grant number 14652016, which is gratefully acknowledged. We would also like to thank Mr. Qiyan Xu for providing the Weibo check-in records data, Dr. Bei Zhao for providing the segmentation code, and Dr. Mingrui Shen for his helpful comments on the initial version of this manuscript.

References

Anselin, L., 1995. Local indicators of spatial association-LISA. Geogr. Anal. 27 (2), 93–115.
Baatz, M., Arini, N., Schäpe, A., Binnig, G., Linssen, B., 2006. Object-oriented image analysis for high content screening: detailed quantification of cells and sub cellular structures with the Cellenger software. Cytometry A 69A (7):652–658. http://dx.doi.org/10.1002/cyto.a.20289.
Bai, X., Shi, P., Liu, Y., 2014. Society: realizing China's urban dream. Nature 509 (7499):158–160. http://dx.doi.org/10.1038/509158a.
Beijing Government, 2005. Beijing Master Urban Plan for 2004–2020, Beijing City.
Chabrier, S., Emile, B., Rosenberger, C., Laurent, H., 2006. Unsupervised performance evaluation of image segmentation. EURASIP J. Adv. Signal Process. 2006:1–13. http://dx.doi.org/10.1155/ASP/2006/96306.
Cheng, Z., Caverlee, J., Lee, K., Sui, D.Z., 2011. Exploring millions of footprints in location sharing services. ICWSM 81–88.
Chongqing Government, 2007. Chongqing Urban and Rural Master Urban Plan for 2007–2020, Chongqing City.
Comaniciu, D., Meer, P., 2002. Mean shift: a robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. 24 (5):603–619. http://dx.doi.org/10.1109/34.1000236.
Dunkel, A., 2015. Visualizing the perceived environment using crowdsourced photo geodata. Landsc. Urban Plan. 142, 173–186.
Elvidge, C.D., Baugh, K.E., Dietz, J.B., Bland, T., Sutton, P.C., Kroehl, H.W., 1999. Radiance calibration of DMSP-OLS low-light imaging data of human settlements. Remote Sens. Environ. 68 (1):77–88. http://dx.doi.org/10.1016/S0034-4257(98)00098-4.
Elvidge, C.D., Tuttle, B.T., Sutton, P.C., Baugh, K.E., Howard, A.T., Milesi, C., ... Nemani, R., 2007. Global distribution and density of constructed impervious surfaces. Sensors 7 (9):1962–1979. http://dx.doi.org/10.3390/s7091962.
Elvidge, C.D., Baugh, K.E., Zhizhin, M., Hsu, F.-C., 2013. Why VIIRS data are superior to DMSP for mapping nighttime lights. Proceedings of the Asia-Pacific Advanced Network. vol. 35, p. 62.
Espindola, G.M., Camara, G., Reis, I.A., Bins, L.S., Monteiro, A.M., 2006. Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. Int. J. Remote Sens. 27 (14), 3035–3040.
Fotheringham, A.S., Brunsdon, C., Charlton, M., 2003. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. John Wiley & Sons.
Frias-Martinez, V., Soto, V., Hohwald, H., Frias-Martinez, E., 2012. Characterizing urban landscapes using geolocated tweets. 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing. IEEE, pp. 239–248.
Garcia-López, M.-À., 2010. Population suburbanization in Barcelona, 1991–2005: is its spatial structure changing? J. Hous. Econ. 19 (2):119–132. http://dx.doi.org/10.1016/j.jhe.2010.04.002.
Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C., 2014. Geo-located Twitter as proxy for global mobility patterns. Cartogr. Geogr. Inf. Sci. 41 (3), 260–271.
Hu, T., Yang, J., Li, X., Gong, P., 2016. Mapping urban land use by using landsat images and open social data. Remote Sens. 8 (2):151. http://dx.doi.org/10.3390/rs8020151.
Jenks, G., 1967. The data model concept in statistical mapping. International Yearbook of Cartography. vol. 7, pp. 186–190.
Jiang, S., Ferreira, J., Gonzalez, M.C., 2012. Discovering urban spatial-temporal structure from human activity patterns. Proceedings of the ACM SIGKDD International Workshop on Urban Computing - UrbComp '12. ACM Press, New York, New York, USA: p. 95 http://dx.doi.org/10.1145/2346496.2346512.
Jiang, B., Ma, D., Yin, J., Sandberg, M., 2016. Spatial, distribution of city tweets and their densities. Geogr. Anal. 48 (3):337–351. http://dx.doi.org/10.1111/gean.12096.
Kaplan, A.M., Haenlein, M., 2010. Users of the world, unite! The challenges and opportunities of social media. Bus. Horiz. 53 (1):59–68. http://dx.doi.org/10.1016/j.bushor.2009.09.003.
Lee, R., Sumiya, K., 2010. Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks - LBSN '10. ACM Press, New York, New York, USA, p. 1.
Li, X., Zhou, Y., 2017. Urban mapping using DMSP/OLS stable night-time light: a review. Int. J. Remote Sens.:1–17 http://dx.doi.org/10.1080/01431161.2016.1274451.
Li, X., Xu, H., Chen, X., Li, C., 2013. Potential of NPP-VIIRS nighttime light imagery for modeling the regional economy of China. Remote Sens. 5 (6):3057–3081. http://dx.doi.org/10.3390/rs5063057.
Liu, X., Wang, J., 2015. The geography of Weibo. Environ. Plan. A 47 (6):1231–1234. http://dx.doi.org/10.1177/0308518X15594912.
Liu, X., Wang, M., 2016. How polycentric is urban China and why? A case study of 318 cities. Landsc. Urban Plan. 151:10–20. http://dx.doi.org/10.1016/j.landurbplan.2016.03.007.
Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., ... Shi, L., 2015. Social sensing: a new approach to understanding our socioeconomic environments. Ann. Assoc. Am. Geogr. 105 (3):512–530. http://dx.doi.org/10.1080/00045608.2015.1018773.
Ma, T., Zhou, C., Pei, T., Haynie, S., Fan, J., 2012. Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: a comparative case study from China's cities. Remote Sens. Environ. 124:99–107. http://dx.doi.org/10.1016/j.rse.2012.04.018.
McMillen, D.P., 2001. Nonparametric employment subcenter identification. J. Urban Econ. 50 (3):448–473. http://dx.doi.org/10.1006/juec.2001.2228.
McMillen, D.P., 2003. The return of centralization to Chicago: using repeat sales to identify changes in house price distance gradients. Reg. Sci. Urban Econ. 33 (3):287–304. http://dx.doi.org/10.1016/S0166-0462(02)00028-5.
McMillen, D.P., 2004. Employment densities, spatial autocorrelation, and subcenters in large metropolitan areas. J. Reg. Sci. 44 (2):225–244. http://dx.doi.org/10.1111/j.0022-4146.2004.00335.x.
McMillen, D.P., McDonald, J.F., 1997. A nonparametric analysis of employment density in a polycentric city. J. Reg. Sci. 37 (4):591–612. http://dx.doi.org/10.1111/0022-4146.00071.
Miller, S.D., Mills, S.P., Elvidge, C.D., Lindsey, D.T., Lee, T.F., Hawkins, J.D., 2012. Suomi satellite brings to light a unique frontier of nighttime environmental sensing capabilities. Proc. Natl. Acad. Sci. 109 (39):15706–15711. http://dx.doi.org/10.1073/pnas.1207034109.
Mitchell, A., 2005. GIS Analysis Volume 2: Spatial Measurements & Statistics. Esri Press.
Ou, J., Liu, X., Li, X., Li, M., Li, W., Parry, M., ... Coltin, K., 2015. Evaluation of NPP-VIIRS nighttime light data for mapping global fossil fuel combustion CO_2 emissions: a comparison with DMSP-OLS nighttime light data. PLoS One 10 (9), e0138310. http://dx.doi.org/10.1371/journal.pone.0138310.
Redfearn, C.L., 2007. The topography of metropolitan employment: identifying centers of employment in a polycentric urban area. J. Urban Econ. 61 (3), 519–541.
Riguelle, F., Thomas, I., Verhetsel, A., 2007. Measuring urban polycentrism: a European case study and its implications. J. Econ. Geogr. 7 (2):193–215. http://dx.doi.org/10.1093/jeg/lbl025.

Shanghai Government, 2001. Shanghai Master Plan 1999–2020. Shanghai People Press, Shanghai (in Chinese).

Shi, K., Yu, B., Huang, Y., Hu, Y., Yin, B., Chen, Z., ... Wu, J., 2014. Evaluating the ability of NPP-VIIRS nighttime light data to estimate the gross domestic product and the electric power consumption of China at multiple scales: a comparison with DMSP-OLS data. Remote Sens. 6 (2):1705–1724. http://dx.doi.org/10.3390/rs6021705.

Small, C., Pozzi, F., Elvidge, C.D., 2005. Spatial analysis of global urban extent from DMSP-OLS night lights. Remote Sens. Environ. 96 (3–4):277–291. http://dx.doi.org/10.1016/j.rse.2005.02.002.

Stefanidis, A., Crooks, A., Radzikowski, J., 2011. Harvesting ambient geospatial information from social media feeds. GeoJournal 78 (2), 319–338.

Steiger, E., Westerholt, R., Resch, B., Zipf, A., 2015. Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. Comput. Environ. Urban. Syst. 54, 255–265.

Sutton, P.C., 2003. A scale-adjusted measure of "urban sprawl" using nighttime satellite imagery. Remote Sens. Environ. 86 (3):353–369. http://dx.doi.org/10.1016/S0034-4257(03)00078-6.

Taubenböck, H., Klotz, M., Wurm, M., Schmieder, J., Wagner, B., Wooster, M., ... Dech, S., 2013. Delineation of central business districts in mega city regions using remotely sensed data. Remote Sens. Environ. 136:386–401. http://dx.doi.org/10.1016/j.rse.2013.05.019.

Weibo Corporation, 2015. Weibo reports third quarter 2015 results. Retrieved from. http://www.nasdaq.com/press-release/weibo-reports-third-quarter-2015-results-20151118-01121.

Yang, J., Song, G., Lin, J., 2015. Measuring spatial structure of China's Megaregions. J. Urban Plann. Dev. 141 (2):4014021. http://dx.doi.org/10.1061/(ASCE)UP.1943-5444.0000207.

Yu, B., Shu, S., Liu, H., Song, W., Wu, J., Wang, L., Chen, Z., 2014. Object-based Spatial Cluster Analysis of Urban Landscape Pattern Using Nighttime Light Satellite Images: A Case Study of China. http://dx.doi.org/10.1080/13658816.2014.922186.

Zagheni, E., Weber, I., 2015. Demographic research with non-representative internet data. Int. J. Manpow. 36 (1), 13–25.

Zhang, Q., Schaaf, C., Seto, K.C., 2013. The vegetation adjusted NTL urban index: a new approach to reduce saturation and increase variation in nighttime luminosity. Remote Sens. Environ. 129:32–41. http://dx.doi.org/10.1016/j.rse.2012.10.022.

Zhong, Y., Zhao, B., Zhang, L., 2014. Multiagent object-based classifier for high spatial resolution imagery. IEEE Trans. Geosci. Remote Sens. 52 (2):841–857. http://dx.doi.org/10.1109/TGRS.2013.2244604.

Zhou, Y., Chen, J., Guo, Q., Cao, R., Zhu, X., 2014. Restoration of information obscured by mountainous shadows through landsat TM/ETM+ images without the use of DEM data: a new method. IEEE Trans. Geosci. Remote Sens. 52 (1):313–328. http://dx.doi.org/10.1109/TGRS.2013.2239651.