Does China's City-Size Distribution Present a Flat Distribution Trend? A Socioeconomic and Spatial Size Analysis From DMSP-OLS Nighttime Light Data

Kaifang Shi, Yizhen Wu, and Shirao Liu

***Abstract*—Inconsistent measurements of city-size and a lack of time-series information on urban socioeconomic development have hindered determining whether China's city-size distribution (CD) follows a Pareto distribution according to multiple perspectives. This article has attempted to evaluate China's CD based on the defense meteorological satellite program- operational line-scan system (DMSP-OLS) nighttime light data in terms of socioeconomic size (SS) and spatial size (SC). First, city size was defined from the DMSP-OLS data. Then, whether China's CD followed a Pareto distribution was evaluated from different perspectives. The results show that China's CD from 1995 to 2015 presents a flat distribution trend; the flat distribution trend of the SC is more obvious than that of the SS; "borrowed size" has become an important reason for the flat trend of China's CD; and residential suburbanization, transportation cost reductions, local government policies, and land finances could effectively explain the CD differences in the flat trends between the SS and SC. This article offers an effective means for quantifying and comparing CD in long time series at a large scale (e.g., national scale or regional scale) and provides a scientific decision basis for governments to build a reasonable CD system in China.**

***Index Terms*—China, city-size distribution (CD), defense meteorological satellite program—operational line-scan system (DMSP-OLS), flat distribution, nighttime light data.**

I. Introduction

CITY-SIZE distribution (CD) has consistently remained a topic of interest in urban geography and economics [1]–[3]. Since Auerbach [4] found that CD followed a Pareto distribution in 1913, an increasing number of scholars have focused on the analysis of this rule [5], [6]. The Pareto distribution indicates that the regression coefficient between logarithmic city ranks and their logarithmic city sizes yields an approximate constant [7], [8]. In other words, the CD presents as a Pareto distribution. Furthermore, Zipf [9] proved that the Pareto distribution coefficient of CD is close to 1 and is typically called Zipf's law. Specifically, Zipf's law implies that in a CD system, the largest city is approximately twice the size of the second city, approximately three times the size of the third city, and so on. If the Pareto distribution coefficient is not equal to 1, it is considered to deviate from Zipf's law [10]. When the coefficient is greater than 1, this indicates that the regional CD is more uniform than that of Zipf, which is commonly called a flat distribution. In contrast, if the coefficient is less than 1, this implies that the CD is more uneven than that of Zipf. Thus, by comparing and analyzing the Pareto distribution coefficient, we can trace and identify the uniformity and rationality of the CD system.

Many studies have analyzed and verified the Pareto CD from different perspectives. Some studies have found that Pareto CD is applicable across different countries or within different countries over time [11]–[13]. For example, Veneri [14] considered the city-size system of the Organisation for Economic Co-operation and Development to follow a Pareto CD. Rosen *et al.* [15] tested cross-sectional samples of global cities and found that CD also followed a Pareto CD. Moreover, Krugman [16] considered that CD exhibited a very good fit of Zipf's law. He also found that many natural elements, such as rivers, followed Zipf's law, which caused considerable debate [17], [18]. Although the Pareto CD has been widely accepted worldwide, many studies have proven that the Pareto coefficient is sensitive to city-size samples [19]. Arshad *et al*. [5] found that the Pareto CD could be effectively supported from an upper-tailed CD. Gabaix [20] also provided a statistical explanation for why the Pareto CD could more effectively reveal the CD of upper-tailed cities. However, Eeckhout [21] indicated that CD should consider upper-tailed cities and also found that CD in the United States followed a lognormal distribution rather than a Pareto CD. In summary, current studies have provided much research on whether CD

Manuscript received February 27, 2021; revised March 31, 2021 and April 15, 2021; accepted May 10, 2021. Date of publication May 12, 2021; date of current version June 11, 2021. This work was supported in part by the MOE (Ministry of Education in China) Project of Humanities and Social Sciences under Grant 18XJC790011, and in part by the Fundamental Research Founds for the Central Universities under Grant XDJK2020B008. *(Corresponding author: Kaifang Shi.)*

The authors are with the School of Geographical Sciences, State Cultivation Base of Ecoagriculture for Southwest Mountainous Land, Southwest University, Chongqing 400715, China, with the Chongqing Jinfo Mountain Field Scientific Observation and Research Station for Kaster Ecosystem, Southwest University, Chongqing 400715, China, and also with the Chongqing Engineering Research Centre for Remote Sensing Big Data Application, School of Geographical Sciences, Southwest University, Chongqing 400715, China (e-mail: shikf1986@126.com; wyz19981013@email.swu.edu.cn; msnlsr@email.swu.edu.cn).

This article has supplementary downloadable material available at https://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2021.3079637

follows Pareto CD or other distribution modes and have reached two consensuses: Pareto CD has been proven by many experimental studies, but there are controversial conclusions; and upper-tailed CD followed by Pareto CD has been widely accepted.

Most of the above studies were concentrated in developed countries [2], [22], which have basically completed their urbanization processes, and their CD is highly stable. Since the 1990s, China has experienced an unprecedented and historic rapid urbanization process. Rapid urbanization has led to the dynamic adjustment of CD [23]–[25]. Due to the large scale of population migration and the particularity of their household registration system, China's CD has unique characteristics [26], [27]. A few studies have attempted to evaluate China's CD and have reached different conclusions. For example, Huang *et al.* [26] indicated that China's CD (e.g., urban land) presents a flat distribution trend from 1992 to 2008. However, Zhao *et al*. [18] found that China's city size fell into the trap of medium-sized city development, which has become increasingly unbalanced. The reasons for these differing conclusions were mainly due to the following issues. First, statistical population data were usually employed to represent city size; however, the different definitions of city population (e.g., *huji population*, *Changzhu population*, or *nonagricultural population*) have led to a diversity of CD results [28]. Second, the adjustment of administrative divisions and the emergence of new cities have created constant changes in city size [29]. Because of the sensitivity of the Pareto coefficient to sample sizes, different distribution results were obtained. Therefore, to attain a more accurate evaluation of the dynamic evolution of CD in China, we need high-quality data that accurately reflects CD to address the problem of inconsistent statistical data.

Traditionally, urban population and urban land (e.g., built-up areas) were used to define city size and then used to quantify and analyze CD [18], [30]. Since urbanization is a complex process, urban population and urban land cannot fully represent the dynamic change of city size. Several studies have indicated that the change in city size caused by urbanization is mainly reflected by the following two factors: socioeconomic size (SS) change and spatial size (SC) change [31]–[33]. The former mainly represents the city size change of the urban population, urban economy, urban culture, and other factors, thus reflecting the change of city social attributes, while the latter mainly shows the spatial scale or structural change of urban land, representing the change of a city's physical structural attributes. Theoretically, the SS and SC are coupled and closely linked. If they are not coordinated, urban sprawl, urban congestion, and other uncoordinated phenomena will result. However, in the previous studies, urban population was widely used as a unilateral representation of SS change, making it difficult to effectively and comprehensively identify the SS distribution law. Additionally, most of the studies individually analyzed SS or SC Pareto distributions, ignoring the differences and connections between them.

Nighttime light data provided by the defense meteorological satellite program—operational line-scan system (DMSP-OLS) have been widely used to characterize human activities [34]–[36], including population [37], gross domestic product (GDP) [38], electricity [39], carbon emissions [40], and urban expansion [41]. These data are regarded as effective means to measure CD [42]–[47]. Due to the inherent characteristics of nighttime light data, the DMSP-OLS data can address the problem of inconsistent statistical data and provide alternative data sources to evaluate SC in a more objective and consistent way. Moreover, the nighttime light intensity can reflect the SS, which has been proven by many studies [48]. The higher the nighttime light intensity is, the more predominant the SS is. Small *et al.* [49] confirmed that the global SC followed a Pareto distribution from the DMSP-OLS data. Taking the DMSP-OLS data as city-size substitutions, Li *et al*. [50] found that China's CD presented a flat distribution trend, but this is often not in line with the reality of urban development to define a natural city with a luminous value greater than 23. It should be noted that the SS and SC were individually evaluated within these studies. Thus, it is necessary to integrate the DMSP-OLS data and other continuous and objective data (e.g., urban land datasets) to simultaneously identify and compare the distributions of SS and SC in China.

To date, the following issues have not been completely resolved.

1) What is the distribution trend of China's city size from the Pareto distribution?
2) What are the similarities and differences of distribution between the SC and SS?

To address the above issues, the SC and SS were defined from DMSP-OLS data. Then, the Pareto CD was tested for selected cities in China, and these cities were further tested within different regions. Third, robust tests were conducted for cities within different provinces and for randomly selected cities. Finally, the driving mechanism of CD was interpreted from institutional and economic perspectives. This study provides a scientific basis for decision-makers to enact reasonable city systems and urban planning policies in China.

II. Study Area and Data Sources

A. Study Area

A total of 236 cities within 30 provincial administrative units (excluding Tibet, Taiwan, Hong Kong, and Macao) in China mainland were selected as experimental objects in this article (see Fig. 1). These cities are evenly distributed in China, and they are cities with concentrated populations and economies. Previous studies usually employed an optical threshold to extract nighttime lights, which were defined as natural cities [10], [50], leading to many small objects that may not be cities in reality. Unlike the above studies, we believe that as the internal social economy is closely linked, cities are regarded as the most basic units for the evaluation of China's socioeconomic environment [51]. Additionally, due to the imbalance of China's socioeconomic development and the hierarchical system of administrative divisions, we divided China's cities into three different regions, including the eastern, central, and western regions, to consider the spatial heterogeneity of Pareto CD. Moreover, cities within some provinces were also chosen to evaluate the robustness of CD.

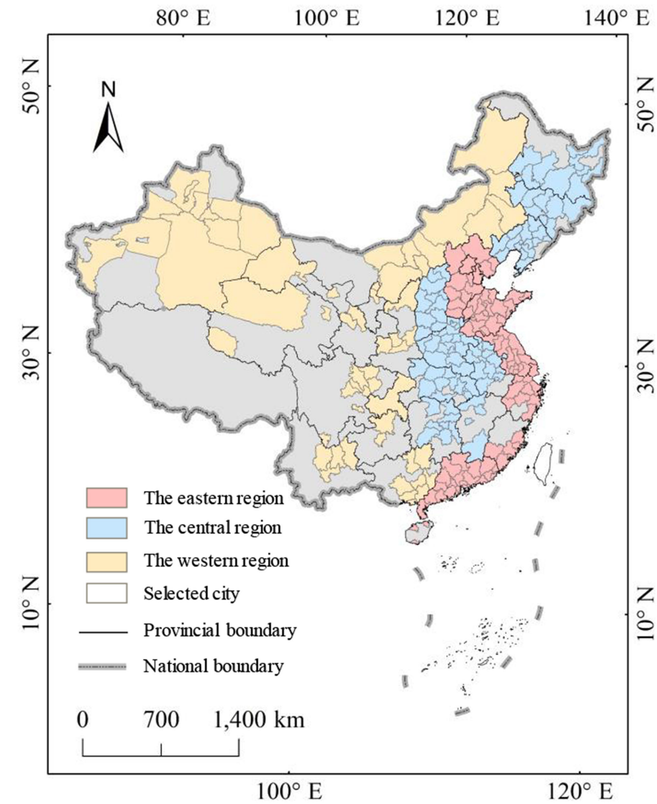


Fig. 1. Spatial distribution of the study area in China. Note: the central and northeastern regions were merged as one region called the central region.

B. Data Sources and Data Preprocessing

The 1995–2013 DMSP-OLS nighttime stable light data were downloaded from the Earth Observation Group.[1] The data were derived from different satellites (F10, F12, F14, F15, F16, and F18) and could detect stable nighttime lights from cities, towns, cars, and other sources while excluding lighting noise such as wildfires, auroras, and other ephemeral events [52]. The digital number value range of the DMSP-OLS data is 0–63, with a spatial resolution of 30 arc seconds, covering most areas of human activities in the world. It is noted that there are two inherent disadvantages of the data: data discontinuities due to derivations from different satellites without onboard calibration; and nighttime light saturation within urban centers due to the limitations of radiation resolution (6-bits). Thus, referring to the studies of Shi *et al.* [39] and Cao *et al*. [53], an invariant region correction was adopted to correct the above defects. Due to the layout limitations, the study will not provide any further elaboration.

Urban land datasets with a spatial resolution of 1-km from 1995 to 2015 were extracted from different remote sensing data and were referenced to the He *et al*. [54] study. The accuracy verification results show that the average overall accuracy was 95.20%, and the average Kappa was 0.66 when compared with the extracted urban land results collected from Landsat TM/ETM+ data. Thus, the datasets provide reliable information for evaluating China's urban expansion.

[1] [Online]. Available: https://payneinstitute.mines.edu/eog/

TABLE I
DETERMINATION COEFFICIENTS (R^2) BETWEEN TOTAL NIGHTTIME LIGHTS (TL) AND POPULATION, GDP, AND EC IN CHINA AT THE CITY LEVEL

Year	R^2		
	TL-Population	TL-GDP	TL-EC
1995	0.6550***	0.8711***	-
2000	0.7551***	0.9092***	0.8747***
2005	0.7223***	0.8856***	0.8140***
2010	0.7328***	0.8551***	0.7801***
2015	0.7434***	0.8426***	0.8070***

Note: Significant at * the 10% level, ** the 5% level, and *** the 1% level.

Administrative vector boundaries for different administrative divisions (cities and provinces) were collected from the National Geomatics Center of China. Finally, all data were projected onto the Albers conic equal area projection with the WGS84 datum, and the spatial data were resampled at a spatial resolution of 1 km.

III. METHODOLOGY

A. Defining City Size

To evaluate CD, the city-size connotation must be defined. Theoretically, the change of city size can be regarded as the result of a 2-D change, including the SC change and SS change [55]. Thus, the DMSP-OLS data and urban land datasets were used to define the SC and SS within the administrative framework. First, urban land was considered the SC in this article. Since the change of SS is a process of population, economy and society, nighttime lights extracted from urban land were thus regarded as the SS, which has been proven by many previous studies [10], [50]. To further verify that the SS can represent city socioeconomic development, the three most representative indicators of urban human activities—population, GDP, and electricity consumption (EC)—were selected to test their correlations at the city level. As given in Table I, there are significant correlations between nighttime lights and these three indicators. Due to the lack of 2015 DMSP-OLS data, 2013 nighttime lights extracted from 2015 urban land were regarded as the 2015 SS in this study. Although the 2015 data were replaced, we believe that this operation does not affect the direction of the final results. Moreover, He *et al.* [56] indicated that urban socioeconomic activities are not only concentrated in urban land but also distributed in the surrounding areas. Thus, nighttime lights extracted from 3-km and 5-km buffered urban land were also used to define the SS to confirm the robustness of the results. It should be noted that the purpose of robustness test is to prevent the uncertainty of variables. Therefore, within a certain threshold range, choosing 3-km or 5-km would not affect the final direction of the results. Also, the 3 km-SS and 5 km-SS have close correlations with urban socioeconomic indicators. In summary, the city size defined in this article not only overcomes the inconsistency of city size but also comprehensively identifies the socioeconomic distribution of city size, which provides an

TABLE II
REGRESSION RESULTS OF CITY SC BASED ON OLS REGRESSION

Year	OLS				
	Top 50	Top 100	Top 150	Top 200	ALL
1995	1.3025	1.1088	1.0326	0.8465	0.4414
2000	1.3723	1.1557	1.1059	0.9950	0.4440
2005	1.5025	1.2028	1.0963	1.0229	0.4768
2010	1.5589	1.2723	1.1281	1.0509	0.4719
2015	1.6786	1.3597	1.2576	1.1587	0.5145
Trend	Flat distribution	Flat distribution	Flat distribution	Flat distribution	Flat distribution

Note: All statistics passed the significance test.

effective and accurate way to evaluate China's city distribution law.

B. Pareto Distribution Analysis

Since the upper tail distribution of city size following the Pareto distribution has been widely accepted, the Pareto distribution is usually given as follows:

$$\mathrm{lnR} = \mathrm{p} - \alpha \mathrm{lnS} + \varepsilon \tag{1}$$

where *p* and α are estimated parameters; *R* and *S* represent city rank and city size, respectively; and ε is the error term. The above formula is generally estimated by the least square (OLS) regression. The test of CD is mainly determined by the Pareto distribution α. When α equals 1, the CD follows the Zipf distribution. If α is greater than 1, then the CD is balanced, showing a flat distribution trend; otherwise, the distribution is relatively concentrated. Gabaix *et al.* [57] noted that OLS regression might cause errors in formula 1, and the following formula was further proposed to verify the Pareto CD

$$\mathrm{ln(R - 1/2)} = \mathrm{p} - \alpha \mathrm{lnS} + \varepsilon. \tag{2}$$

Ultimately, both regressions were used to verify China's CD within different regions in this study. Five years (1995, 2000, 2005, 2010, and 2015) were treated as validated years in this study because the city-size system generally does not change significantly in a short time interval.

IV. RESULTS AND DISCUSSION

A. Distribution of the City-Size System in China

According to the Pareto regression, all R^2 values are greater than 0.7, suggesting that Pareto's law can well fit the CD in China. Because the Pareto regression is sensitive to city samples, the top 50, 100, 150, 200, and all cities were selected for the regression test. As given in Table II, all Pareto coefficients (α) of the SC are significant at the 1% significance level. From a horizontal perspective, the Pareto coefficients gradually decreased with the increase in city samples. The results imply that large cities are more evenly distributed than that of small and medium-sized cities in China. With the increase of the year from 1995 to 2015, the Pareto coefficients of the SC also increase. For example, the coefficients increased from 1.303 in 1995 to 1.679 in 2015 for the top 50 cities. Similar results were also found in the top 100, 150, 200, and all cities. This means that China's SC distribution presents a flat distribution trend. Table III gives the results of the Pareto regression of the SS. From the perspectives of changes of time and city samples, the coefficients show an increasing trend, which also indicates that the SS presents a flat distribution trend. To prevent a deviation of results, the Gabaix–Ibragimov regression was also conducted for the SS and SC. Similar results were concluded in Tables S1 and S2. Moreover, the 3-km and 5-km buffered SS were also verified in this article, and the coefficients presented a growth trend with the decrease of sample size and the update of time (see Tables S3–S6).

Compared with the SC and SS, we found that both present a flat distribution trend (see Tables II and III). With the decrease in city samples and time changes, the flat distribution trend is increasingly obvious. Interestingly, when taking all cities as samples, the Pareto coefficients are significantly less than 1, which shows that China's urban development is significantly unbalanced, and urban land expansion and socioeconomic development are mainly concentrated in some large cities. Additionally, the flat distribution trend of the SC is more obvious and faster than that of the SS from 1995 to 2015. On the one hand, due to the rapid development of China, a large number of real estate developments and new-area construction have led to the rapid expansion of urban land in large, medium-sized, and small cities [58]. On the other hand, due to more robust medical treatment, education, and infrastructure in large cities, the development speed in large cities is often faster than that of small and medium-sized cities [18]. This indirectly implies that urban land expansion is faster than urban socioeconomic development, which may partly lead to urban sprawl and ghost city phenomena, which have been proven in many previous studies [48], [59]. Although the results show that the CD tends to be flat, we found that the Pareto coefficients of the SS are 0.160–0.275 from 1995 to 2015, indicating that China's urban socioeconomic development is extremely unbalanced. For example, some provinces have implemented the strategy of strengthening provincial capital, which has led to great differences in city development within the provinces.

B. Distribution of City-Size Systems Within Different Regions

Considering the regional imbalance of China's development, the Pareto CD was evaluated in the eastern, central, and western regions. The top 20 and 40 cities were selected to estimate the

TABLE III
REGRESSION RESULTS OF CITY SS BASED ON OLS REGRESSION

Year	OLS				
	Top 50	Top 100	Top 150	Top 200	ALL
1995	1.2264	1.0418	0.9503	0.8002	0.1603
2000	1.2770	1.0830	1.0221	0.9104	0.1887
2005	1.4075	1.1257	1.0119	0.9332	0.2349
2010	1.4924	1.1958	1.0631	0.9748	0.2356
2015	1.6158	1.2788	1.1631	1.0657	0.2751
Trend	Flat distribution	Flat distribution	Flat distribution	Flat distribution	Flat distribution

Note: All statistics passed the significance test.

TABLE IV
REGRESSION RESULTS OF CITY SC FOR THE TOP 20 CITIES AT THE REGIONAL SCALE

Year	Eastern region		Central region		Western region	
	OLS	Gabaix-Ibragimov	OLS	Gabaix-Ibragimov	OLS	Gabaix-Ibragimov
1995	1.0891	1.2339	1.1891	1.3595	1.1389	1.2994
2000	1.1168	1.2677	1.2258	1.4048	1.2170	1.3793
2005	1.4029	1.5959	1.2540	1.4382	1.3249	1.5098
2010	1.4866	1.6897	1.3044	1.4985	1.3201	1.5149
2015	1.5823	1.7995	1.4115	1.6230	1.2795	1.4589
Trend	Flat distribution	Flat distribution	Flat distribution	Flat distribution	Flat distribution	Flat distribution

Note: All statistics passed the significance test.

TABLE V
REGRESSION RESULTS OF CITY SS FOR THE TOP 20 CITIES AT THE REGIONAL SCALE

Year	Eastern region		Central region		Western region	
	OLS	Gabaix-Ibragimov	OLS	Gabaix-Ibragimov	OLS	Gabaix-Ibragimov
1995	1.0391	1.1814	1.1522	1.3103	0.9454	1.0770
2000	1.0399	1.1820	1.1199	1.2753	1.0344	1.1733
2005	1.2781	1.4537	1.1297	1.2860	1.1904	1.3575
2010	1.4131	1.6070	1.1587	1.3188	1.2429	1.4206
2015	1.4967	1.7028	1.2245	1.3988	1.2462	1.4215
Trend	Flat distribution	Flat distribution	Flat distribution	Flat distribution	Flat distribution	Flat distribution

Note: All statistics passed the significance test.

Pareto coefficients based on formulas 1 and 2, respectively (see Tables IV and V; and Tables S7–S8). The results show that the CD presents a flat trend within different regions. For the SC within the top 20 cities, the Pareto coefficients estimated by the eastern region are larger than those of other regions. This implies that the CD within large cities is flatter in the eastern region. This is also in line with the reality of China. Many large cities are distributed in the eastern region, including Beijing, Shanghai, Guangzhou, Shenzhen, Nanjing, Tianjin, and Hangzhou. The Pareto coefficients in the western region declined slightly after 2010. One possible reason is the result of repaid urban sprawl in some western provincial capitals. Similar results were found in the top 40 cities (see Tables S7 and S8). For the SS within the top 20 cities, urban development in the eastern region is still the most balanced, and the Pareto coefficients show a rapid growth trend. An interesting phenomenon is that the central region has the largest Pareto coefficients of the SS within the top 40 cities. This might be the eastern industrial transfer undertaken by the central region, and many cities have made great progress. Furthermore, the Pareto coefficients of the 3-km and 5-km buffered SS were also estimated in this article (see Tables S9–S12) and demonstrated that China's SS presents a flat distribution within different regions.

Compared with different regions within the top 20 cities, the SC of the central region was significantly higher than that of the SS (see Tables IV and V). Although the results show that the CD within the central region is more balanced, they also imply that the top 20 cities have experienced rapid urban land expansion from 1995 to 2015. When considering the top 40 cities, this trend has not changed, indirectly indicating that there is an imbalance between urban socioeconomic development and urban land expansion in the central region (see Tables S7 and S8). Additionally, the Pareto coefficients of the SC are greater than those of the SS within the eastern region, but the trend is lower than that in the central region. The similar development trends of the SS and SC in the western region show that the CD is more balanced, which also indirectly shows a slow urban development trend. In summary, the CD within different regions presents an obvious flat trend, and the SC distribution is more even than that of the SS.

TABLE VI
REGRESSION RESULTS OF CITY SIZE FOR RANDOMLY SELECTED CITIES WITHIN THE TOP 100 CITIES IN CHINA

Type		1995	2000	2005	2010	2015	Trend
City spatial size	OLS (Top 20)	0.8583	0.7938	1.0777	1.2646	1.2942	Flat distribution
	Gabaix-Ibragimov (Top 20)	0.9849	0.9142	1.2286	1.4792	1.4784	Flat distribution
	OLS (Top 40)	0.9974	1.0906	1.1903	1.1432	1.3031	Flat distribution
	Gabaix-Ibragimov (Top 40)	1.0908	1.2035	1.3546	1.2749	1.4814	Flat distribution
City socioeconomic size	OLS (Top 20)	0.9718	1.135	0.8072	1.5979	1.2503	Flat distribution
	Gabaix-Ibragimov (Top 20)	1.1436	1.34	0.9272	1.7984	1.4596	Flat distribution
	OLS (Top 40)	0.9134	0.9033	1.1070	1.1130	1.1624	Flat distribution
	Gabaix-Ibragimov (Top 40)	1.0107	0.9945	1.2132	1.2129	1.2879	Flat distribution

Note: All statistics passed the significance test.

C. Comparison of the Capacity to Estimate the GDP and EPC

Many studies have shown that administrative division and economic development are important factors affecting the estimation of the Pareto coefficient [42]. To further test the robustness of the above results, two methods were selected for validation. First, we randomly selected cities within the top 100 cities every year in China. The top 20 and 40 cities within the top 100 cities were further chosen to evaluate the Pareto CD. Because these randomly selected cities could exclude administrative or economic associations, we could examine whether the random city-size samples would affect the Pareto coefficient estimation. Second, due to the influence of institutional and administrative means, the inner cities within the province are closely related. Thus, the provincial internal cities could be employed to test the upper tail distribution of the Pareto estimation. Due to the limitations of city number, the top 20 cities within a province were selected for estimation. Finally, six provinces, including Anhui, Shandong, Henan, Guangdong, Sichuan, and Yunnan, met the evaluation conditions in this study.

The regression results of randomly selected cities within the top 100 cities are given in Table VI. Regardless of whether the top 20 or 40 cities are selected for estimation with different regression models, the Pareto coefficients show a growth trend. This means that even considering the administrative and economic ties between cities, the CD still shows a flat trend, that is, it deviates from the Zipf's law. Compared with the Pareto coefficients between the SS and SC, the SC coefficients are larger than those of the SS, showing a flat trend. Similar results can be concluded in Table S13.

Fig. 2 presents the Pareto coefficients of the top 20 cities within different provinces from different regression models. The CD of all provinces shows a flat distribution trend, which is basically consistent with the above results. Additionally, two interesting phenomena are shown in Fig. 2. First, the Pareto coefficients of the SS and SC are approximately 1 or greater than 1 within Shandong and Henan, suggesting that the CD in these two provinces is reasonable. Second, most of the estimated coefficients are significantly less than 1 in Guangdong, Sichuan, and Yunnan. The reason may be that education, medical treatment, industry, business, population, and urban land expansion are concentrated in a few large cities within these provinces. Fig.

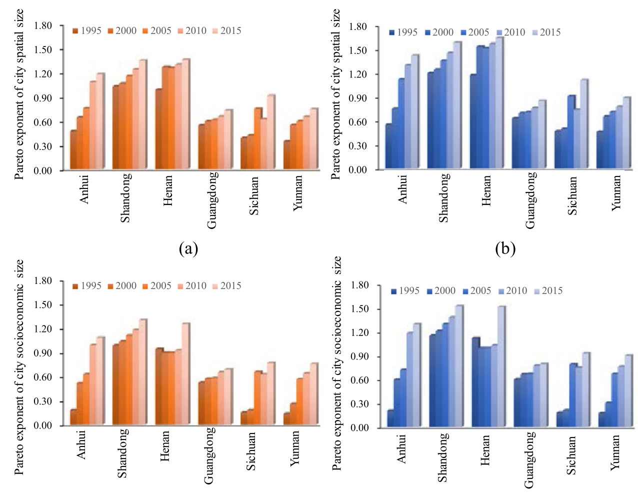


Fig. 2. Regression results of city size for the top 20 cities within different provinces. (a) and (c) Operational line-scan system. (b) and (d) Gabaix–Ibragimov. Note: All statistics passed the significance test.

S1 also presents similar results and further verifies the credibility of the estimation.

Through the above analysis, we can quickly and effectively conclude that China's CD presents a flat trend from the DMSP-OLS data and urban land datasets. This article can solve two significant difficulties for the traditional data: the inconsistent measurement of city size and a lack of spatiotemporal information on urban socioeconomic development. To further verify the robustness of the results, a comparison analysis with previous studies was also evaluated in this article. For the SC, Huang *et al*. [26], and Li *et al.* [50] found similar increasing trends for the Pareto coefficients based on multisource data in China from 1995 to 2010. For the SS, many studies have demonstrated that the Pareto coefficients present an increased tendency when using the population to represent city size [60]. In addition, further verification can be provided with the DMSP/OLS data, and Huang *et al*. [26] found that the SC Pareto coefficients increased from 0.79 in 1995 to 0.82 in 2000. Although there are some differences between the above results and the results of this article, this is due to the differences in size samples and data sources.

D. Driving Mechanism of CD Change

Through the above analysis, we found that China's CD, including SC and SS, does not meet Zipf's law. The estimated Pareto coefficients increased from 1995 to 2015, highlighting a flat trend of CD in China. Why has this flat distribution trend been formed? It is necessary to identify the driving mechanism of CD in China. Most studies have attributed this flat trend to the effects of institutional and strategic forces, including the following two views. The first view is that due to the restriction of China's household registration system, the labor flow between different cities is not sufficient, which hinders the agglomeration of population to large cities and then leads to a flat trend of CD [26]. With the gradual opening of China's household registration system, a large number of people have moved freely to developed areas, but this flat trend has not changed. This implies that the household registration system is not the main reason for the flat trend. The second view indicates that the strategy of restricting large cities is an important reason for the flat trend of China's CD. However, the Chinese government has also put forward the strategy of "coordinating development of large, medium, and small cities." Additionally, some provinces have proposed the strategy of strengthening provincial capitals and focusing on the development of major cities such as provincial capitals. The above phenomena cannot absolutely explain the flat trend of CD.

Some studies have attempted to explain the flat trend from economic views. Fan *et al.* [61] noted that high housing prices in large cities would restrict the flow of population, which would lead to the flat trend of China's CD. However, the reality of high housing prices did not completely limit the flow of population. When considering city-size growth in the United States, Alonso [62] found that small cities close to large cities usually grew faster than small cities not close to large cities. He called this phenomenon "the borrowed size." Specifically, small cities could obtain the benefits of agglomeration economies and avoid negative externalities from large cities. Therefore, these small cities could lead to the diversion of the economy and population from large cities, and this phenomenon is irreversible, resulting in a flat trend of CD. Taking Kunshan, Zhangjiagang, and Changshu as examples, due to their proximity to Shanghai, their borrowed size was formed from Shanghai, which showed some economic characteristics similar to Shanghai. Specifically, from 2000 to 2010, the population size of these three cities increased rapidly, and the urban population increased from 434 400, 432 300, and 565 200 to 1 118 600, 762 600, and 929 100, respectively, [63]. Additionally, they have avoided the high rent and housing prices caused by the scarcity of land resources, which led to the social and economic development level being much higher than other small and medium-sized cities. The above phenomenon may be an important cause for the flat trend of China's CD. Meanwhile, with the improvement of transportation infrastructure, especially the construction of expressways and high-speed railway networks, an increasing number of small and medium-sized cities could make use of the advantages of a large city agglomeration economy to form development advantages while avoiding uneconomical agglomeration. For example, Deng *et al*. [64] indicated that China's high-speed rail has realized the close connection between different cities and promoted the rapid development of medium-sized and small cities. Ma *et al*. [58] found that during 2000–2010, 51% of the newly added urban population in China related to the county (city) region, with an urban population growth of about 80 million, accounting for 27.6% of the country's total from 22.2%. Thus, we believe the above evaluation can effectively explain the flat trend of China's CD.

In sections 3.1–3.3, we note that China's SC distribution is flatter than its SS distribution. This indirectly shows that China's cities have experienced rapid urban land expansion [65], especially in medium-sized and small cities. Moreover, the shortening gap trend of the SC between different cities is obviously larger than that of the SS from 1995 to 2015. Thus, it is necessary to analyze the driving mechanism of this phenomenon. Four aspects could explain the above phenomena. First, residential suburbanization has accelerated urban sprawl [66], but there is no corresponding rapid growth of urban social and economic development. Due to the limitation of land resources in large cities, residential suburbanization mainly took place in medium-sized and small cities, which leads to a more balanced distribution of the SC [67]. Second, the reduction of transportation cost is also an important driving force for the spread of urban land expansion [64]. Third, under the GDP-oriented performance appraisal mechanism, local governments actively lead urban expansion through urban planning guidance, large-scale project development, and infrastructure construction. Although the social economy has been developed to a certain extent, this reflects rapid urban land expansion. In this case, the distribution of the SC appears to be flatter than that of SS [68]. Fourth, the dependence of local governments on land finance is another key factor leading to large-scale urban land expansion, which also increases the risk of urban socioeconomic development [69].

V. Conclusion

This article has attempted to evaluate China's CD based on the DMSP-OLS data in terms of SS and SC perspectives. First, city size was defined. Then, whether China's CD followed a Pareto distribution was evaluated. Two important phenomena were found in this article: China's SC and SS distributions present a flat distribution trend. The flat distribution trend of the SC is more obvious and faster than that of the SS. Moreover, the heterogeneity and robustness of the results were verified by different regions. The driving mechanism of CD was discussed, and we indicated that "the borrowed size" may be an important reason for the flat trend of CD. Additionally, residential suburbanization, transportation cost reduction, local government policy, and land finance can effectively explain why the SC distribution in China is flatter than SS distribution. This article provides a scientific decision basis to build a reasonable CD system for the Chinese government. The combination of the DMSP-OLS data and urban land datasets can be used to quantify the CD from the regional scale to the national scale, which provides an effective means for us to study and compare CD in different regions and countries.

References

[1] G. Anderson and Y. Ge, "The size distribution of Chinese cities," *Regional Sci. Urban Econ.,* vol. 35, no. 6, pp. 756–776, Nov. 2005.

[2] B. Jiang and T. Jia, "Zipf's law for all the natural cities in the united states: A geospatial perspective," *Int. J. Geographical Inf. Sci.,* vol. 25, no. 8, pp. 1269–1281, Aug. 2011.

[3] S. Zhao *et al.*, "Contemporary evolution and scaling of 32 major cities in China," *Ecolog. Appl.,* vol. 28, no. 6, pp. 1655–1668, Sep. 2018.

[4] F. Auerbach, "Das gesetz der bevlkerungskonzentration," *Petermanns Geographische Mitteilungen,* vol. 49, pp. 74–76, 1913.

[5] S. Arshad, S. Hu, and B. N. Ashraf, "Zipf's law and city size distribution: A survey of the literature and future research agenda," *Physica , Statist. Mech. Appl.,* vol. 492, pp. 75–92, Feb. 2018.

[6] A. Ermias, J. Bogaert, and F. Wogayehu, "Analysis of city size distribution in Ethiopia: Empirical evidence from 1984 to 2012," *J. Urban Manage.,* vol. 8, no. 2, pp. 237–244, Aug. 2019.

[7] B. J. L. Berry and A. Okulicz-Kozaryn, "The city size distribution debate: Resolution for US urban regions and megalopolitan areas," *Cities,* vol. 29, no. S1, pp. 17–23, Mar. 2012.

[8] M. Bee, M. Riccaboni, and S. Schiavo, "Distribution of city size: Gibrat, Pareto, Zipf," in *Proc. Math. Urban Morphol.*, 2019, pp. 77–91.

[9] G. K. Zipf, "Human behavior and the principle of least effort," *Amer. J. Sociol.,* vol. 110, p. 306, 1949.

[10] B. Jiang, J. Yin, and Q. Liu, "Zipf's law for all the natural cities around the world," *Int. J. Geographical Inf. Syst.,* vol. 29, no. 3-4, pp. 498–522, Feb. 2015.

[11] Y. Chen, "The rank-size scaling law and entropy-maximizing principle," *Physica A, Statsit. Mech. Appl.,* vol. 391, no. 3, pp. 767–778, Feb. 2012.

[12] Y. Wu, M. Jiang, Z. Chang, Y. Li, and K. Shi, "Does China's urban development satisfy Zipf's law? A multiscale perspective from the NPP-VIIRS nighttime light data," *Int. J. Environ. Res. Public Health,* vol. 17, no. 4, Feb. 2020, Art. no. 1460.

[13] R. Lagonigro, J. C. Martori, and P. Apparicio, "Understanding airbnb spatial distribution in a southern European city: The case of Barcelona," *Appl. Geogr.,* vol. 115, Feb. 2020, Art. no. 102136.

[14] P. Veneri, "City size distribution across the OECD: Does the definition of cities matter?," *Comput., Environ. Urban Syst.,* vol. 59, pp. 86–94, Sep. 2016.

[15] R. Kenneth and R. Mitchel, "The size distribution of cities: An examination of the Pareto law and primacy," *J. Urban Econ.,* vol. 6, pp. 165–186, Sep. 1980.

[16] K. Paul, "Confronting the mystery of urban hierarchy," *J. Jpn. Int. Economies,* vol. 10, pp. 399–418, Dec. 1996.

[17] G. Wan, D. Zhu, C. Wang, and X. Zhang, "The size distribution of cities in China: Evolution of urban system and deviations from Zipf's law," *Ecolog. Indicators,* vol. 111, Apr. 2020, Art. no. 106003.

[18] S. Zhao, D. Zhou, C. Zhu, Y. Sun, and S. Liu, "Spatial and temporal dimensions of urban expansion in China," *Environ. Sci. Technol.,* vol. 49, no. 16, pp. 9600–9609, Aug. 2015.

[19] J. Luckstead and S. Devadoss, "Do the world's largest cities follow Zipf's and Gibrat's laws?," *Econ. Lett.,* vol. 125, no. 2, pp. 182–186, Nov. 2014.

[20] X. Gabaix, "Zipf's law for cities: An explanation," *Quart. J. Econ.*, vol. 114, pp. 739–767, Aug. 1999.

[21] J. Eeckhout, "Gibrats law for (all) cities," *Amer. Econ. Rev.,* vol. 94, no. 5, pp. 1429–1451, Dec. 2004.

[22] M. Modica, "The impact of the European union integration on the city size distribution of the member states," *Habitat Int.,* vol. 70, pp. 103–113, Dec. 2017.

[23] Z. Zhu *et al.*, "Understanding an urbanizing planet: Strategic directions for remote sensing," *Remote Sens. Environ.,* vol. 228, pp. 164–182, Jul. 2019.

[24] X. Guan, H. Wei, S. Lu, Q. Dai, and H. Su, "Assessment on the urbanization strategy in China: Achievements, challenges and reflections," *Habitat Int.,* vol. 71, pp. 97–109, Jan. 2018.

[25] F. Xu, Z. Wang, G. Chi, and Z. Zhang, "The impacts of population and agglomeration development on land use intensity: New evidence behind urbanization in China," *Land Use Policy,* vol. 95, Jun. 2020, Art. no. 104639.

[26] Q. Huang *et al.*, "Detecting the 20 year city-size dynamics in China with a rank clock approach and DMSP/OLS nighttime data," *Landscape Urban Plan.,* vol. 137, pp. 138–148, May 2015.

[27] Y. Wang, Y. Liu, Y. Li, and T. Li, "The spatio-temporal patterns of urban–rural development transformation in China since 1990," *Habitat Int.,* vol. 53, no. 53, pp. 178–187, Apr. 2016.

[28] Z. Xu and Z. Nong, "City size distribution in China: Are large cities dominant?," *Urban Stud.,* vol. 46, no. 10, pp. 2159–2185, Aug. 2009.

[29] L. Zhang, J. Lei, X. M. Li, and C. Gao, "The features and influencing factors of urban expansion in China during 1997–2007," *Prog. Geography,* vol. 30, no. 5, pp. 607–614, 2011.

[30] X. Peng, "China's demographic history and future challenges," *Science,* vol. 333, no. 6042, pp. 581–587, Jul. 2011.

[31] J. Chen, L. Zhuo, P.-J. Shi, and I. Toshiaki, "The study on urbanization process in China based on DMSP/OLS data: Development of a light index for urbanization level estimation," *J. Remote Sens.,* vol. 7, no. 3, pp. 168–175, Jan. 2003.

[32] S. Shao, X. Li, and J. Cao, "Urbanization promotion and haze pollution governance in China (in English)," *Econ. Res. J.,* no. 2, pp. 148–165, Feb. 2019.

[33] K. Shi, J. Shen, L. Wang, M. Ma, and Y. Cui, "A multiscale analysis of the effect of urban expansion on $pM_{2.5}$ concentrations in China: Evidence from multisource remote sensing and statistical data," *Building Environ.,* vol. 174, May 2020, Art. no. 106778.

[34] N. Levin *et al.*, "Remote sensing of night lights: A review and an outlook for the future," *Remote Sens. Environ.,* vol. 237, Feb. 2020, Art. no. 111443.

[35] L. Zhang, X. Li, and F. Chen, "Spatiotemporal analysis of venezuela's nighttime light during the socioeconomic crisis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 13, pp. 2396–2408, May 2020.

[36] M. M. Bennett and L. C. Smith, "Advances in using multitemporal night-time lights satellite imagery to detect, estimate, and monitor socioeconomic dynamics," *Remote Sens. Environ.,* vol. 192, pp. 176–197, Apr. 2017.

[37] M. Tan *et al.*, "Modeling population density based on nighttime light images and land use data in China," *Appl. Geogr.,* vol. 90, no. 46, pp. 239–247, Jan. 2018.

[38] K. Shi *et al.*, "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.,* vol. 6, no. 2, pp. 1705–1724, Jan. 2014.

[39] K. Shi *et al.*, "Detecting spatiotemporal dynamics of global electric power consumption using DMSP-OLS nighttime stable light data," *Appl. Energy,* vol. 184, pp. 450–463, 2016.

[40] K. Shi *et al.*, "Modeling spatiotemporal CO_2 (carbon dioxide) emission dynamics in China from DMSP-OLS nighttime stable light data using panel data analysis," *Appl. Energy*, vol. 168, pp. 523–533, Dec. 2016.

[41] T.-H. K. Chen, A. V. Prishchepov, R. Fensholt, and C. E. Sabel, "Detecting and monitoring long-term landslides in urbanized areas with nighttime light data and multi-seasonal landsat imagery across Taiwan from 1998 to 2017," *Remote Sens. Environ.,* vol. 225, pp. 317–327, May 2019.

[42] Z. Deng, M. Qin, and S. Song, "Re-study on Chinese city size and policy formation," *China Econ. Rev.,* vol. 60, 2020, Art. no. 101390.

[43] R. Ch, D. A. Martin, and J. F. Vargas, "Measuring the size and growth of cities using nighttime light," *J. Urban Econ.,* no. 6, Apr. 2020, Art. no. 103254.

[44] J. Gibson and C. Li, "Pareto's law and city size in China: Diverging patterns in land and people," in *Population Change and Impacts in Asia and the Pacific*, J. Poot and M. Roskruge, Eds., New York, NY, USA: Springer, 2020, pp. 29–47.

[45] X. Li, X. Li, D. Li, X. He, and M. Jendryke, "A preliminary investigation of Luojia-1 night-time light imagery," *Remote Sens. Lett.*, vol. 10, no. 6, pp. 526–535, Jun. 2019.

[46] X. Li *et al.*, "Anisotropic characteristic of artificial light at night–Systematic investigation with VIIRS DNB multi-temporal observations," *Remote Sens. Environ.,* vol. 233, Nov. 2019, Art. no. 111357.

[47] X. Li, N. Levin, J. Xie, and D. Li, "Monitoring hourly night-time light by an unmanned aerial vehicle and its implications to satellite remote sensing," *Remote Sens. Environ.,* vol. 247, Sep. 2020, Art. no. 111942.

[48] Z. Chen, B. Yu, Y. Hu, C. Huang, K. Shi, and J. Wu, "Estimating house vacancy rate in metropolitan areas using NPP-VIIRS nighttime light composite data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 8, no. 5, pp. 2188–2197, May 2015.

[49] C. Small, F. Pozzi, and C. D. Elvidge, "Spatial analysis of global urban extent from DMSP-OLS night lights," *Remote Sens. Environ.,* vol. 96, no. 3, pp. 277–291, Jun. 2005.

[50] S. Li and X. Liu, "The flat distribution of city size in China:A multidimensional regional validation and its economic interpretations," (in Chinese), *J. World Economy,* no. 11, pp. 144–169, 2017.

[51] Y. Su *et al.*, "China's 19-year city-level carbon emissions of energy consumptions, driving forces and regionalized mitigation guidelines," *Renew. Sustain. Energy Rev.,* vol. 35, pp. 231–243, Jul. 2014.

[52] C. D. Elvidge *et al.*, "A fifteen year record of global natural gas flaring derived from satellite data," *Energies,* vol. 2, no. 3, pp. 595–622, Sep. 2009.

[53] Z. Cao, Z. Wu, Y. Kuang, and N. Huang, "Correction of DMSP-OLS nightime light images and its application in China," (in Chinese), *J. Geo-Inf. Sci.,* vol. 17, no. 9, pp. 1092–1102, Jan. 2015.

[54] C. He, Z. Liu, J. Tian, and Q. Ma, "Urban expansion dynamics and natural habitat loss in China: A multiscale landscape perspective," *Global Change Biol.,* vol. 20, no. 9, pp. 2886–2902, Sep. 2014.

[55] C. Yang *et al.*, "A spatial-socioeconomic urban development status curve from NPP-VIIRS nighttime light data," *Remote Sens.,* vol. 11, no. 20, Oct. 2019.

[56] C. He, B. Gao, Q. Huang, Q. Ma, and Y. Dou, "Environmental degradation in the urban areas of China: Evidence from multi-source remote sensing data," *Remote Sens. Environ.,* vol. 193, pp. 65–75, May 2017.

[57] X. Gabaix and R. Ibragimov, "Rank − 1/2: A simple way to improve the OLS estimation of tail exponents," *J. Bus. Econ. Statist.,* vol. 29, no. 1, pp. 24–39, 2011.

[58] Q. Ma, C. He, and J. Wu, "Behind the rapid expansion of urban impervious surfaces in China: Major influencing factors revealed by a hierarchical multiscale analysis," *Land Use Policy,* vol. 59, pp. 434–445, Dec. 2016.

[59] B. Gao, Q. Huang, C. He, Z. Sun, and D. Zhang, "How does sprawl differ across cities in China? A multi-scale investigation using nighttime light and census data," *Landscape Urban Plan.,* vol. 148, no. 41, pp. 89–98, Apr. 2016.

[60] Y. Chen and Y. Zhou, "Scaling laws and indications of self-organized criticality in urban systems," *Chaos, Solitons Fractals,* vol. 35, no. 1, pp. 85–98, Jan. 2008.

[61] J. Fan and T. Shao, "Housing price, location of diversified product and urban system," (in Chinese), *Econ. Res. J.,* no. 2, pp. 87–99, 2011.

[62] W. Alonso, "Urban zero population growth," *Daedalus,* vol. 102, pp. 191–206, 1973.

[63] X. Li and D. Zheng, "The characteristics of urbanization and the formation of urban system," (in Chinese), *Urban Planning Forum*, vol. 233, no. 1, pp. 19–29, 2017.

[64] T. Deng, C. Gan, A. Perl, and D. Wang, "What caused differential impacts on high-speed railway station area development? Evidence from global nighttime light data," *Cities,* vol. 97, Feb. 2020, Art. no. 102568.

[65] A. Schneider and C. Mertes, "Expansion and growth in chinese cities, 1978–2010," *Environ. Res. Lett.,* vol. 9, no. 2, Feb. 2014, Art. no. 024008.

[66] L. Lu, H. Guo, C. Corbane, and Q. Li, "Urban sprawl in provincial capital cities in China: Evidence from multi-temporal urban land products using landsat data," *Sci. Bull.,* vol. 64, no. 14, pp. 955–957, Jul. 2019.

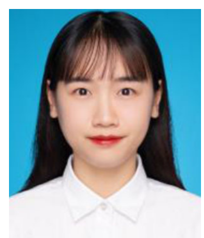
[67] J. Chen and Y. Liu, "Review of the causes and mechanism of urban sprawl in China," (in Chinese), *Modern Urban Res.,* no. 7, pp. 31–36, 2018.

[68] C. Ding, "Policy and praxis of land acquisition in China," *Land Use Policy,* vol. 24, no. 1, pp. 1–13, Jan. 2007.

[69] T. Zhang, "Land market forces and government's role in sprawl: The case of China," *Cities,* vol. 17, no. 2, pp. 123–135, Apr. 2000.

Kaifang Shi received the Ph.D. degree in cartography and geographic information systems from East China Normal University, Shanghai, China, in 2017.

He is currently an Associate Professor with the School of Geographical Sciences, Southwest University, Chongqing, China. His research interests include remote sensing analysis in socioeconomic characterization and urbanization process, nighttime light remote sensing, and urban geography.



Yizhen Wu is currently working toward the M.S. degree in cartography and geographic information science with the School of Geographical Sciences, Southwest University, Chongqing, China.

Her research interests include urban remote sensing, nighttime light remote sensing, and remote sensing applications.

Shirao Liu is currently working toward the M.S. degree incartography and geographic information system at the School of Geographic Sciences, Southwest University, Chongqing, China.

Her research interests include nighttime light remote sensing and urban geography.