remote sensing

Article

Monitoring the Distribution and Variations of City Size Based on Night-Time Light Remote Sensing: A Case Study in the Yangtze River Delta of China

Yuan Ding [1,†], Jia Hu [1,†], Yingbao Yang [1,*], Wenyu Ma [1], Songxiu Jiang [1], Xin Pan [1], Yong Zhang [1,2], Jingjing Zhu [1] and Kai Cao [3,4]

1 School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China; dingyuanhhu@hhu.edu.cn (Y.D.); jiahu@hhu.edu.cn (J.H.); mawy@hhu.edu.cn (W.M.); 181309040002@hhu.edu.cn (S.J.); px1013@hhu.edu.cn (X.P.); zyrs@hhu.edu.cn (Y.Z.); 181309030007@hhu.edu.cn (J.Z.)
2 College of Geography and Planning, Chizhou University, Chizhou 247000, China
3 School of Geographic Sciences, East China Normal University, Shanghai 200241, China; kcao@geo.ecnu.edu.cn
4 Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai 200241, China
* Correspondence: yyb@hhu.edu.cn
† These authors contributed equally to this work.

Abstract: Effectively monitoring the size of a city in real time enables the scientific planning of urban development. Models that utilize the distribution and variations in city size generally use population data as inputs, which cannot be obtained in a timely and rapid manner. However, night-time light (NTL) remote sensing may be an alternative method. A case study was carried out on the Yangtze River Delta (YRD) in China, and the rank–size rule, the law of primate cities, and the Gini coefficient were employed to monitor the variation in city size in the study area. The urban areas extracted based on NTL remote sensing were utilized instead of the traditionally used population data to evaluate the variations in city size from 2012 to 2017. Considering the empiricism and subjectivity of the thresholding method, urban areas were extracted from NTL data combined with the normalized differential vegetation index and land-surface temperature data based on the artificial neural network algorithm. Based on the results, the YRD did not fit the distribution of the primate cities from 2012 to 2017. However, this region satisfied the rank–size rule well, which indicated that the development of medium–small cities was more prominent than that of larger cities, and the dispersed force was larger than the concentrated force. Notably, the city size reached a relatively balanced level in the study area. Further, sensitivity analysis revealed that the relatively low extraction accuracy of urban areas of few small cities had little effect on the results of city size variations. Moreover, the validation of city size computed from statistical population data and its comparison with results calculated based on the statistical data of urban areas aligned with the results of this study, which indicates the rationality and applicability of monitoring the variations in city size using the urban areas extracted from NTL remote sensing instead of population data.

Keywords: night-time light data; extraction of urban areas; rank–size rule; law of primate city

check for updates

Citation: Ding, Y.; Hu, J.; Yang, Y.; Ma, W.; Jiang, S.; Pan, X.; Zhang, Y.; Zhu, J.; Cao, K. Monitoring the Distribution and Variations of City Size Based on Night-Time Light Remote Sensing: A Case Study in the Yangtze River Delta of China. *Remote Sens.* **2022**, *14*, 3403. https://doi.org/10.3390/rs14143403

Academic Editors: Ran Goldblatt, Steven Louis Rubinyi and Hogeun Park

Received: 16 June 2022
Accepted: 14 July 2022
Published: 15 July 2022

1. Introduction

Dramatic urbanization has occurred globally, and is mainly due to rapid economic development and population growth [1]. According to statistical data, more than 54% of land is populated by humans worldwide. In fact, the total population in urban areas is predicted to exceed 2 billion by 2050 [2]. Rapid urbanization can lead to several environmental issues, including air pollution, urban heat islands, climate change, shortage of resources, and pressure for sustainability worldwide [3–6]. Rapid urbanization may also result in unhealthy development in urban systems, especially developing countries. For instance, the development of big cities was better than that of small and medium cities. Accordingly, the development of city size in an urban system was recognized to be unbalanced. A comprehensive understanding of variations in city size is thus useful for urban planning and decision making, and monitoring the distribution and variations in city size, which is also helpful for understanding the development model and rational planning of urban systems, is of great significance.

Several popular measurement methods for analyzing the distribution and variations in city size include the rank–size rule [7], the law of the primate city [8], and the Gini coefficient [9]. Briefly, the urban population or urban area is input into the above models [10], and the distribution and variation trends of city size are analyzed based on the calculated indices. Although the most direct and effective monitoring method is population data, there are some problems with this method [11–13]. First, the statistical standards for population data vary among countries. For instance, population statistics in China are mainly based on administrative units, as urban and rural areas are divided according to the administrative boundaries of urban districts and towns. Socioeconomic units in other countries are generally based on functional areas, such as metropolitan areas in the United States. Moreover, considerable time and labor are required to obtain population data. Night-time light (NTL) remote sensing, which has emerged in recent years, has enabled new directions for research on city monitoring, as it can provide real-time information about the earth with the same observation standard. Several studies employed remote-sensing data to monitor the urbanization of cities and assess socioeconomic activities [14–16]. Nitsch [17] showed that urban area and population have similar distribution trends when measuring urban size distribution. Therefore, in this study, urban areas extracted from NTL remote sensing were used to replace the traditionally used population data in these city size analysis models to monitor the distribution and variations in city size.

Many satellite images of different resolutions can be used to extract urban areas for cities that are suitable for different scenarios. High- and medium-high-resolution images, such as Landsat series [18], are suitable for exploring urbanization at the city scale, but are not suitable for large scales, such as urban and global levels. Mapping urban areas on a large scale requires several high-resolution images with cloud-free images, which also requires significant time and labor for processing. Coarse-resolution satellite images, such as MODIS and NTL data, are suitable for large-scale urban area extraction [1,19,20]. However, NTL data can be utilized to monitor human activities, as it captures city light at night, which separates urban areas from the surrounding suburbs [21,22]. According to prior studies, the NTL value has a highly close connection with population, gross domestic product, and built-up areas [23,24]. Thus, this value is more suitable for selecting NTL data for the analysis of variations in city size compared to other optimal satellite images.

Currently, commonly used NTL data are obtained from two NTL satellites, the Defense Meteorological Satellite Program-Operational Linescan System (DMSP-OLS) and Suomi National Polar-Orbiting Partnership Visible Infrared Imaging Radiometer Suite (NPP-VIIRS). NPP-VIIRS has better quality than DMSP-OLS, as it decreases the problem of the saturation effect that exists in DMSP-OLS [25–27]. Thus, NPP-VIIRS is better choice for extracting urban built-up areas. Recently, methods for extracting urban areas based on NPP-VIIRS data mainly include two types: thresholding and classification. The thresholding method has been used by many researchers, as it is simple and easy to conduct [28–32]. Nevertheless, this method is too subjective and empirical to determine the thresholds for extracting urban areas, owing to considerable uncertainty among cities at different development levels [33]. To resolve such limitations of the thresholding method, some researchers have utilized machine learning methods for the extraction of urban areas. For instance, Jing et al. [34] applied four machine learning algorithms to extract urban areas of eastern China. Based on their results, the machine learning methods achieved good accuracy and lower sensitivity than the thresholding method. Xu et al. [35] extracted urban areas from NTL data by utilizing an artificial neural network algorithm, which led to a more accurate quality result than the thresholding method. Hence, extracting urban areas based on machine learning is better than extraction achieved with the traditional thresholding approach. The selection of training samples is another technical problem in urban area extraction using a classification method based on machine learning. Extracting training samples from NTL data alone may lead to the overestimation or underestimation of urban areas. To improve the extraction of urban areas, some researchers have added other remote-sensing data for sample selection, including normalized differential vegetation index (NDVI) and land-surface temperature (LST) data [34,35]. Notably, a better extraction result was achieved when the NTL data were combined with multi-source remote-sensing data to select training samples to extract urban areas based on the machine learning method.

The objectives of this study were to extract urban built-up areas in the study region using a combination of NTL data and NDVI and LST data based on the machine learning algorithm, and to monitor the distribution and variations in city size in the study area by utilizing the models of rank–size rule, the law of primate city, and the Gini coefficient, using extracted urban areas instead of the traditional population data.

2. Methodology

2.1. Study Area

The Yangtze River Delta (YRD) is located in eastern China and is adjacent to the East China Sea (Figure 1). This region has a subtropical monsoon climate, with an annual mean temperature of 16.9 °C. The region also has three provinces, namely Jiangsu, Zhejiang, and Anhui, and one international city, Shanghai city, with a total of 26 cities. The YRD covers an area of 217,700 km^2 and has a population of 150 million. Owing to its unique geographical location and natural resources, the YRD urban agglomeration has become the most developed and largest economic zone in China, and it is one of six metropolitan agglomerations in the world. Further, this area has experienced tremendous urban expansion over the past few decades.

Figure 1. Location of the YRD in China and demarcation of the study area.

2.2. Data Collection and Preprocessing

Multiple datasets were used in this study (Table 1), including NTL data and four MODIS datasets. Version 1 NPP-VIIRS NTL data, on a monthly basis from 2012 to 2017, were obtained from the National Oceanic and Atmospheric Administration (NOAA) (https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html (accessed on 11 November 2020)). Additionally, the four MODIS data from the same time series, which were 16-day Normalized Differential Vegetation Index (NDVI) product (MOD13A1), 8-day MODIS land-surface temperature (LST) product (MOD11A2), annual global land water mask product (MOD 44w), and annual MODIS Land Cover Type product (MCD12Q1), were obtained from National Aeronautics and Space Administration (NASA) (https://ladsweb.modaps.eosdis.nasa.gov/search/ (accessed on 17 November 2020)). The NPP-VIIRS, MOD 13A1

NDVI, and MOD 11A2 LST data were used to select samples for urban area extraction for the study area. MCD12Q1 land-cover data were used to determine the optimal combination of coefficients for urban area extraction. The MOD 44w data were employed to remove the light intensity of water bodies from the NPP-VIIRS images. In this study, all remote-sensing data were reprojected to the Albers Conical Equal Area projection and resampled to a resolution of 500 m for consistency with the NTL and land-cover data. The annual NPP-VIIRS and LST data were reproduced by the mean value composition in this study, while the annual NDVI data were reproduced based on the maximum value composition to decrease the influence of cloud contamination [36]. The population and actual urban area data from 2012 to 2017 were obtained from the statistical yearbooks of provinces and cities. The population data were utilized to calculate the indices for city size analysis, and the statistical urban areas were used for comparison with urban areas extracted using our NTL remote-sensing-based method. In this study, the permanent urban population at the end of the year was regarded as the urban population of each city in the study area.

Table 1. Data used in this study.

Data	Description	Resolution	Time
NPP VIIRS	Night-time light	500 m/Month	
MOD 13A1	Normalized Differential Vegetation Index	500 m/16 Days	2012–2017
MOD 11A2	Land-surface temperature	1 km/8 Days	
MOD 44w	Global land water mask	250 m/Year	2012–2017
MCD 12Q1	Land-cover type	500 m/Year	2012–2017
Population		Year	2012–2017
Statistical urban areas		0.01 km^2	2012–2017

2.3. Monitoring of the Distribution and Variations in City Size

In this study, the rank–size rule, the law of the primate city, and the Gini coefficient were employed to analyze the distribution and variations in city size for the study area. The rank–size rule is usually represented by Zipf's index, and the law of the primate city is calculated using the urban primacy index. Zipf's index, the urban primacy index, and the Gini coefficient were calculated from the acreage of the extracted urban areas. The three indices were also computed from the population data to validate the results obtained from the extracted urban areas. If the variation trend of the three indices calculated from both data sources is the same, urban areas extracted from NTL remote sensing can replace the population to analyze the variation in the differences in city size.

2.3.1. Rank–Size Rule

The rank–size rule was utilized to reveal the size distribution in an urban system based on the relationship between the size of cities and the rank of city sizes, which can also reflect the development level of a country or an urban system. In 1949, Zipf provided a theoretical foundation [7], which can be expressed as follows:

$$\mathrm{lg}P_i = \mathrm{lg}P_1 - q\mathrm{lg}R_i \quad (1)$$

where lg is the basic logarithm with base 10, R_i is the rank of city i according to city size among all cities, P_1 is the city size of the largest city, P_i is the city size of city i, and q is Zipf's index.

The ideal Zipf index should be 1, indicating balanced urban development within the urban system. However, in general, Zipf's index is not equal to one. An index greater than 1 indicates that the development of big cities is prominent, while that of small and medium-sized cities is underdeveloped. On the contrary, an index lower than 1 indicates that the development of big cities is not prominent, whereas that of small- and medium-sized cities is better. An increasing trend in Zipf's index indicates that the concentrated force in the distribution of city size is larger than the dispersed force, which means that the development of city size in large cities is faster than that in small cities. Conversely, a decrease in Zipf's index indicates that the dispersed force in the urban system is larger than the concentrated force, and the development of city size in small cities is faster than in big cities.

2.3.2. Law of Primate City

The urban primacy index was utilized to describe the extent of concentration of the urban population in the primate city in an urban system, which can largely reflect the development features of the urban system. This index was proposed by Mark [8], which indicates that the size of the largest city is markedly greater than that of other cities in an urban system. The index is equal to the ratio of the population of the largest city to that of the second largest city in the study area, which can be described by the following formula:

$$S_2 = P_1/P_2 \tag{2}$$

where S_2 is a 2-city index and P_1 and P_2 are the populations of the largest and second-largest cities, respectively.

According to some researchers, a 2-city index is too one-sided. Thus, the 4-city index and 11-city index were proposed, and can be calculated as follows:

$$S_4 = P_1/(P_2 + P_3 + P_4) \tag{3}$$

$$S_{11} = 2P1/\sum_{i=2}^{11} Pi \tag{4}$$

where S_4 and S_{11} are the 4-city index and 11-city indices, respectively, P_1 is the population of the largest city in the urban system, and P_2 to P_{11} is the population of the second to the eleventh city based on city size.

The threshold of the 2-city index should be 2, while the thresholds of the 4-city and 11-city indices should be 1. The greater the number of cities involved in the calculation of the urban primacy index, the more reliable the results. Accordingly, the 11-city index was selected to compute the urban primacy index for the study area. In this study, the urban primacy index was calculated from the age of the extracted urban areas and the population in the study area.

2.3.3. Gini Coefficient

The Gini coefficient was originally used to measure the income gap between residents in a country or region, as proposed by Gini in 1912 [9]. Marshall applied the Gini model to cities to analyze city development based on differences in size by analyzing the degree of population aggregation on the whole urban agglomeration, and proposed the concept of the city Gini coefficient for the first time, which can be calculated as follows:

$$G = T/2S(n-1) \tag{5}$$

where n is the total number of cities in the urban system, S is the total population of these cities in the urban system, and T is the sum of the absolute differences in population size among these cities in the urban system.

As an indicator of the relative difference, the Gini coefficient can reflect the balance degree of the population distribution in various cities, ranging from 0 to 1. The closer the Gini coefficient to the value of 0, the more dispersed the city size, while the closer the coefficient to the value of 1, the more concentrated the city size. The value range of the Gini coefficient can be divided into five groups: values below 0.2, 0.2, 0.3, 0.3, 0.4, 0.4, 0.5, and greater than 0.5. These groups represent the absolute balance of city size, relative balance of city size, relative reasonable city size, significant difference in city size, and abnormal disparity, respectively.

2.4. Urban Area Extraction

In this study, NPP-VIIRS NTL, NDVI, and LST data were utilized to extract the urban areas of the study area based on an artificial neural network (ANN) algorithm. The main steps for extracting urban areas are described below. First, the pixels occupied by water bodies were removed using global land water mask data, and potential urban areas were regarded as pixels with NTL values greater than 1, according to Yang et al. [37].

The thresholds of the three datasets utilized to select samples were subsequently determined based on the standard deviation from the mean value. Notably, urban areas were extracted city by city in the study area, as the development and economic levels markedly vary in the study area, and the accuracy of the extracted results may be low if urban areas of different cities are extracted as a whole. The equations are as follows:

$$\begin{cases} T_j^{VIIRS} = Mean_j^{VIIRS} + aStd_j^{VIIRS} \\ T_j^{NDVI} = Mean_j^{NDVI} + bStd_j^{NDVI} \\ T_j^{LST} = Mean_j^{LST} + cStd_j^{NDVI} \end{cases} \quad (6)$$

where T_j^{VIIRS}, T_j^{NDVI}, and T_j^{LST} are the thresholds of the NPP-VIIRS, NDVI, and LST data for city j, respectively; $Mean_j^{VIIRS}$, $Mean_j^{NDVI}$, and $Mean_j^{LST}$ are the mean values of the three datasets for city j; Std_j^{VIIRS}, Std_j^{NDVI}, and Std_j^{NDVI} are the standard deviations of these data for city j; and a, b, and c are coefficients, which are normally set to −1, −0.5, 0, 0.5, and 1, respectively. Different values of T_j^{VIIRS}, T_j^{NDVI}, and T_j^{LST} can be obtained according to the values of a, b, and c.

Based on the features of high NTL values, high temperatures, and low vegetation coverage in urban areas, the training samples of urban and non-urban areas were selected based on the following formulas:

$$Sample(i,j) = \begin{cases} VIIRS(i,j) > T_j^{VIIRS} \& NDVI(i,j) < T_j^{NDVI} \& LST(i,j) > T_j^{LST}, Urban \\ VIIRS(i,j) < T_j^{VIIRS} \& NDVI(i,j) > T_j^{NDVI} \& LST(i,j) < T_j^{LST}, Non-urban \end{cases} \quad (7)$$

where $Sample(i,j)$ represents the selected samples of pixel i of city j, and $VIIRS(i,j)$, $NDVI(i,j)$, and $LST(i,j)$ are the pixel values of the NPP-VIIRS, NDVI, and LST data for pixel i of city j, respectively.

Therefore, there are 125 combinations of coefficients a, b, and c. The optimal set of combinations was selected based on the following standard:

$$\min|S_k - S_0| \& \max\{g\_mean_k\} \quad (8)$$

where S_k and S_0 are the number of total pixels of the extracted urban areas based on the NTL data and the number of total pixels of the urban areas in the land-cover data, respectively, and g_mean_k is the geometric mean precision, which is introduced below.

Based on the optimal set of coefficients confirmed above, the optimal thresholds for the three datasets used to select samples for urban area extraction were determined in the third step. Thereafter, the samples were selected and randomly divided into approximately 10% training samples and 90% test samples. An ANN algorithm was employed to extract the urban areas of the study area based on the training samples.

ANN is a type of machine learning model that has been widely utilized in many studies owing to its self-adapting, self-organizing, and self-learning advantages. A feedforward network structure composed of an input layer, a hidden layer, and an output layer was utilized in this study [35]. The input layer consists of three nodes: NTL, LST, and NDVI. Based on a sensitivity analysis, the number of hidden nodes was set to five. The output nodes have two possible results: urban and non-urban areas. The pixels that met the expectations of urban and non-urban areas were outputted as (1, 0) and (0, 0), respectively.

After training, the test samples were utilized to validate the accuracy of urban area extraction for the study area. In this study, in addition to the Kappa coefficient used, the geometric mean accuracy was also used [38]. The geometric mean accuracy was calculated as follows:

$$g\_mean = \sqrt{\frac{d}{m+d} \times \frac{d}{n+d}} \quad (9)$$

where g_mean is the geometric mean accuracy, d is the number of matched urban area pixels between NTL data and land-cover data, m is the number of urban area pixels for which NTL data do not match the land-cover data, and n is the number of urban area pixels for which the land-cover data do not match the NTL data.

3. Results and Discussion

3.1. Urban Expansion of the YRD

3.1.1. Accuracy Assessment of Urban Area Extraction

In this study, the kappa coefficient and geometric mean accuracy were used to validate the accuracy of the urban area extraction of the YRD. Table 2 shows the average geometric mean accuracy and kappa coefficient for the YRD. As depicted in the table, the urban area extraction of YRD based on the method proposed in this study achieved high accuracy, with an average geometric mean accuracy and kappa coefficient ranging from 79.21% to 84.88% and 0.761 to 0.828, respectively. The geometric mean accuracy of all years was between 82% and 85%, except for 2015, which had a value of 79.21%. In 2012 and 2015, the Kappa coefficients were 0.793 and 0.761; however, in other years, this value was approximately 0.82. Figures 2 and 3 show the distribution of geometric mean accuracy and Kappa coefficient of urban area extraction for all cities in the YRD from 2012 to 2017. This result indicates that most cities in the YRD have a high geometric mean accuracy and Kappa coefficient for six years, with values above 90% and 0.80, respectively. However, the accuracy of several medium- and small-sized cities, such as Anqing, Chizhou, Xuancheng, and Chuzhou of the Anhui province, which had relatively lower economic development, was relatively low, with a geometric mean accuracy and Kappa coefficient of less than 60% and 0.60, respectively.

3.1.2. Spatial–Temporal Variations of Urban Areas

Figure 4 displays the distribution of the extracted urban areas of the study area from 2012 to 2017 and highlights an obvious spatial discrepancy in urban areas. This finding indicates that most urban areas within the study area were distributed in international metropolises (Shanghai, China), metropolitan circles (Suzhou, Changzhou, and Wuxi, China), and provincial capitals (Nanjing, Ningbo, Hangzhou, and Hefei, China) during the six years. On the contrary, only a small proportion of urban areas were scattered in medium- and small-sized cities. As shown in Figure 5, the acreage of all cities in the study area increased gradually from 2012 to 2017. The urban areas of all cities in the study area extended along the central urban areas to the surrounding suburbs (Figure 4). However, the increase in the urban areas of big cities was small, such as Shanghai, Suzhou, Wuxi, and Nanjing, the urban areas of which increased from 2012 to 2014. In contrast, the extension force of the urban areas of medium- and small-sized cities, such as Xuancheng, Shaoxing, and Nantong, increased in recent years, especially from 2015 to 2017.

Table 2. Mean accuracy of urban area extraction in the YRD.

Accuracy	2012	2013	2014	2015	2016	2017
g-mean (%)	82.10	84.88	83.06	79.21	83.94	84.10
Kappa	0.793	0.828	0.805	0.761	0.816	0.817

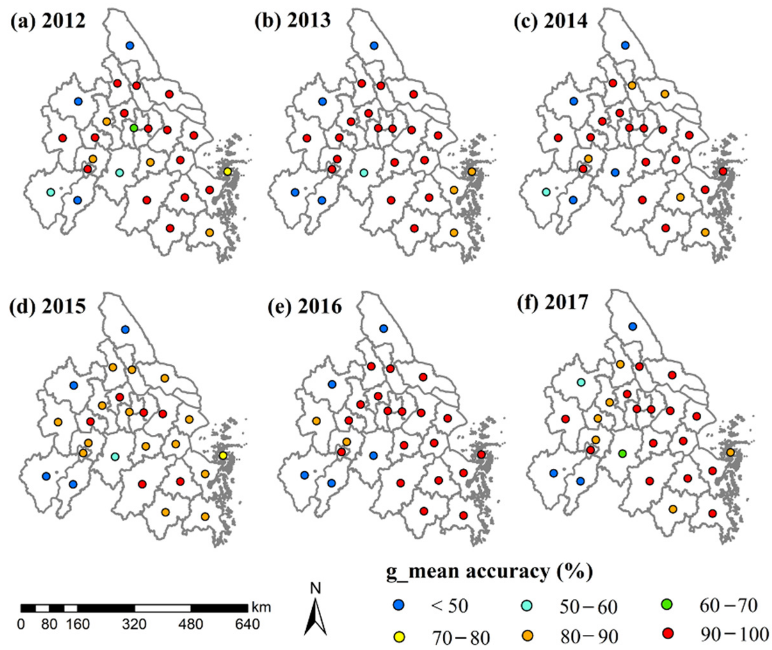


Figure 2. Geometric mean accuracy for urban area extraction in the YRD.

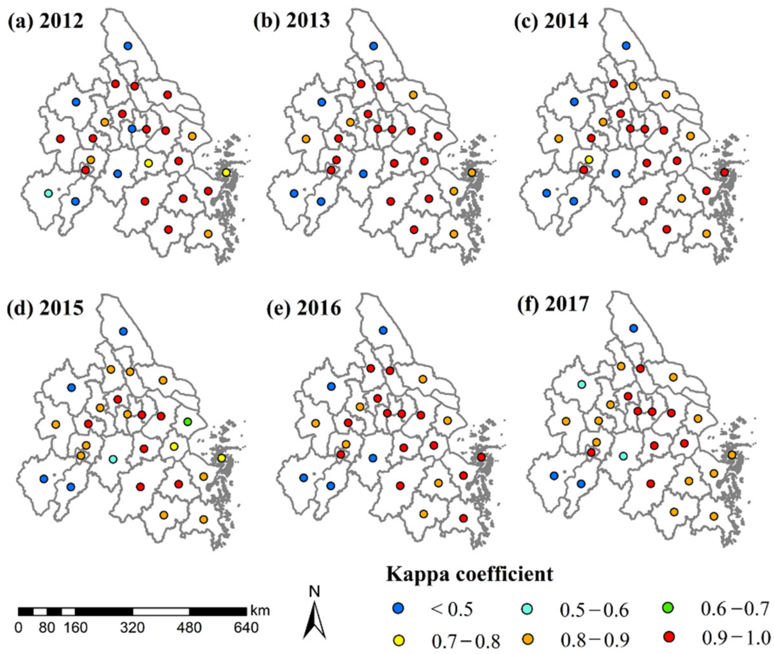


Figure 3. Kappa coefficient for urban area extraction in the YRD.

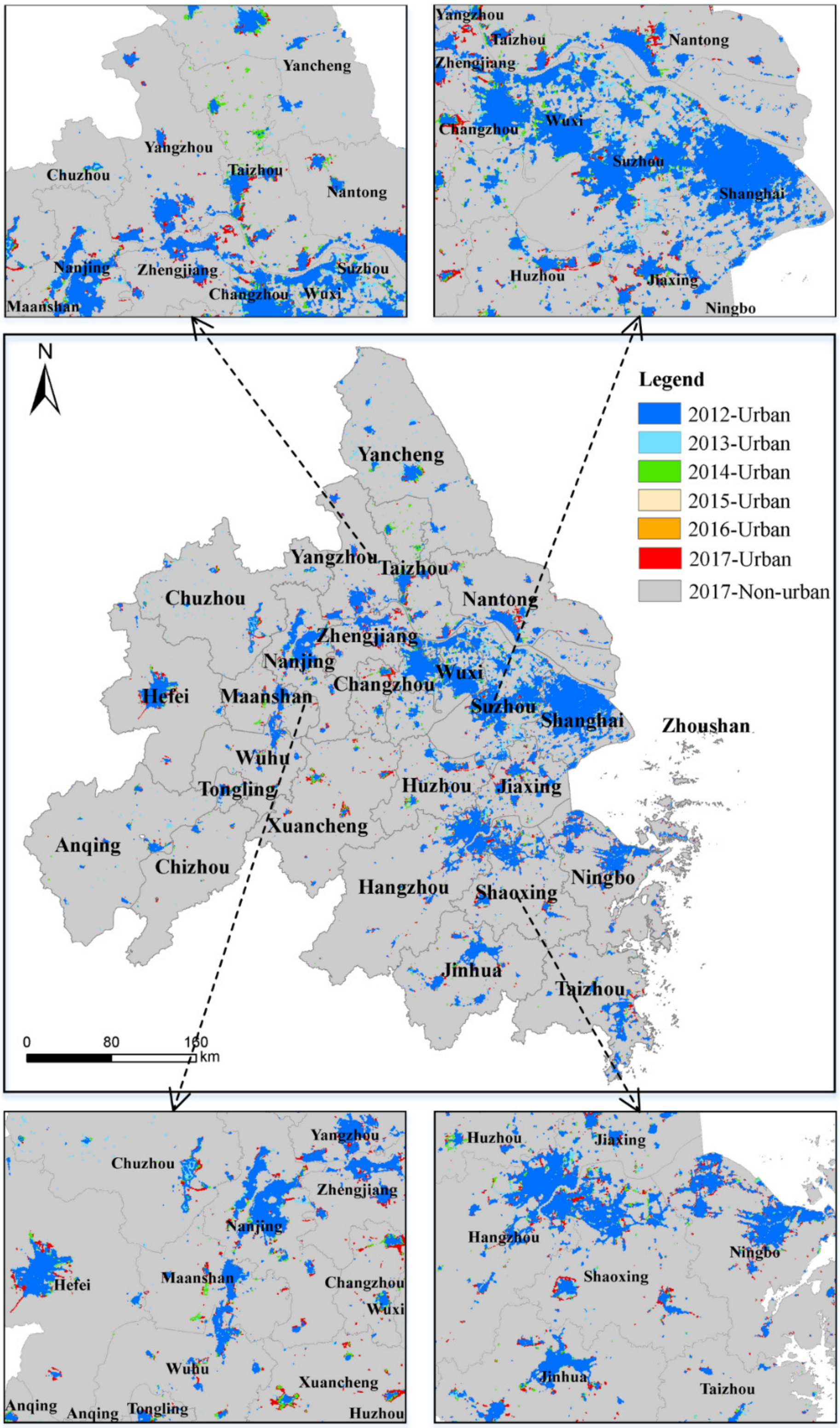


Figure 4. Urban expansion of the study area from 2012 to 2017.

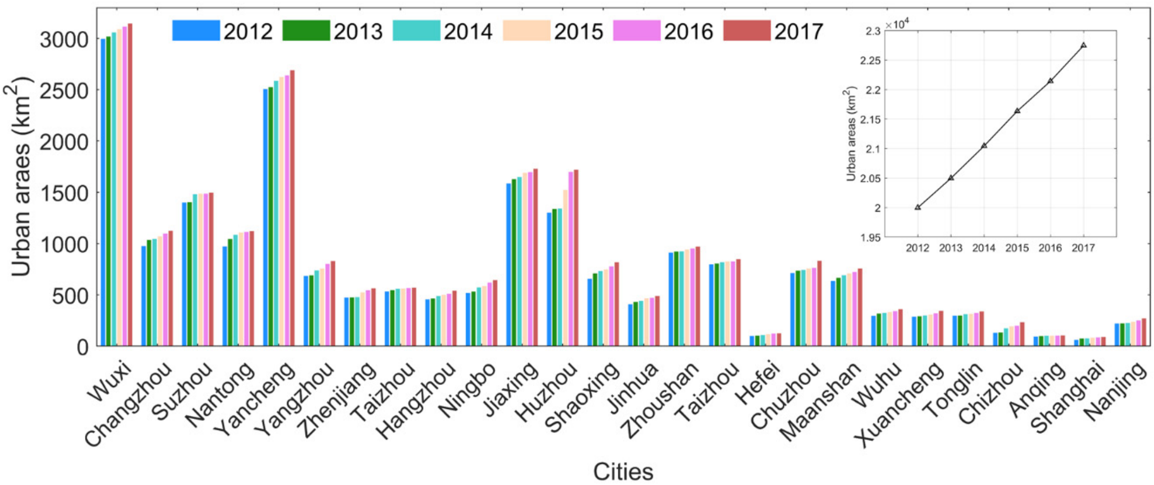


Figure 5. The increase in urban areas in all cities in the study area from 2012 to 2017.

3.2. Variations in City Size in the YRD

3.2.1. Variations in Rank–Size

Table 3 shows the fitting results of the double logarithm of the rank–size model in the YRD from 2012 to 2017. These results indicate that both the urban areas extracted from remote-sensing data and the population data can fit the model well, with R^2 values of approximately 0.79 and greater than 0.83 from 2012 to 2017, respectively. Although the value of R^2 fitted from urban areas was lower than that fitted from population size, the value met the requirements of the rank–size model. Therefore, a subsequent analysis of the variations in rank size of the YRD from 2012 to 2017 was performed.

Table 3. Fitting results of rank–size in the YRD from 2012 to 2017 (y represents Zipf's index; x in the left and right columns represent urban area and urban population, respectively).

Year	Urban Areas		Population	
	Fitting Equations	R^2	Fitting Equations	R^2
2012	y = −1.048x + 3.782	0.796	y = −0.884x + 3.347	0.840
2013	y = −1.034x + 3.784	0.801	y = −0.880x + 3.352	0.838
2014	y = −1.019x + 3.784	0.795	y = −0.875x + 3.357	0.835
2015	y = −1.008x + 3.788	0.798	y = −0.856x + 3.351	0.860
2016	y = −1.010x + 3.793	0.797	y = −0.852x + 3.357	0.860
2017	y = −0.982x + 3.791	0.789	y = −0.845x + 3.361	0.858

Zipf's index was calculated from urban areas and the population size of the YRD from 2012 to 2017, respectively; the values are shown in Figure 6. The Zipf's index calculated from the extracted urban areas from 2012 to 2017 was below the value of one, which indicates that the development of big cities in the YRD was not prominent, while that of small- and medium-sized cities was better. In addition, a decreasing Zipf's index trend calculated from the extracted urban areas was found, with a value of 0.704 in 2012, which decreased to 0.651 in 2017. Such findings indicate that the dispersed force in the YRD was larger than the concentrated force. Thus, the development of the city size of small cities is faster than that of big cities in the YRD. However, from 2010 to 2013, the urban area in the Yangtze River Delta expanded rapidly, while the resident population increased slowly, especially in small and medium-sized cities. Thus, the Zipf's index calculated from urban area is bigger than that calculated by population.

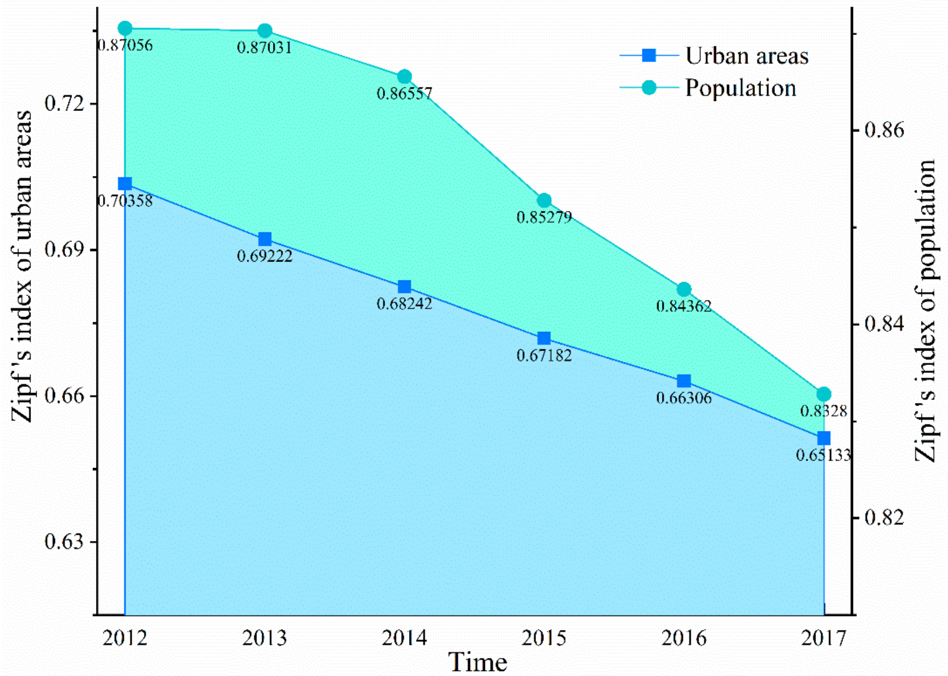


Figure 6. Variations in Zipf's index for the extracted urban areas and the statistical population data of the YRD from 2012 to 2017.

The Zipf's index calculated from the population data of the YRD displayed the same distribution as that calculated from the extracted urban areas, which indicates the feasibility of analyzing the variation in rank–size by utilizing NTL remote sensing.

3.2.2. Variations in Primate City

In this study, the 11-city index of urban primacy was utilized to analyze the variations in urban primacy in the study area. Figure 7 shows the distribution of the YRD urban primacy index from 2012 to 2017. As depicted in the figure, the urban primacy index calculated from the extracted urban areas of the YRD was lower than one from 2012 to 2017, which indicates that the YRD agglomeration did not belong to the distribution of the primate city during this period. Additionally, the urban primacy index of the study area decreased gradually, declining from 0.506 in 2012 to 0.471 in 2017. Such a decrease indicates that the size distribution of the primate city in the YRD economic region had a relatively decreasing trend compared with other cities, and the status and function of the primate city in the study area gradually reduced during this period. Such findings indirectly illustrate that the size of medium- and small-sized cities in the study area increased gradually during this period.

The urban primacy index computed from population data also displayed a similar distribution and variation trend to that calculated from the extracted urban areas, which was also below the value of one and decreased gradually from 2012 to 2017. This finding indicates that the application of NTL remote sensing for the analysis of variations in urban primacy is a feasible method.

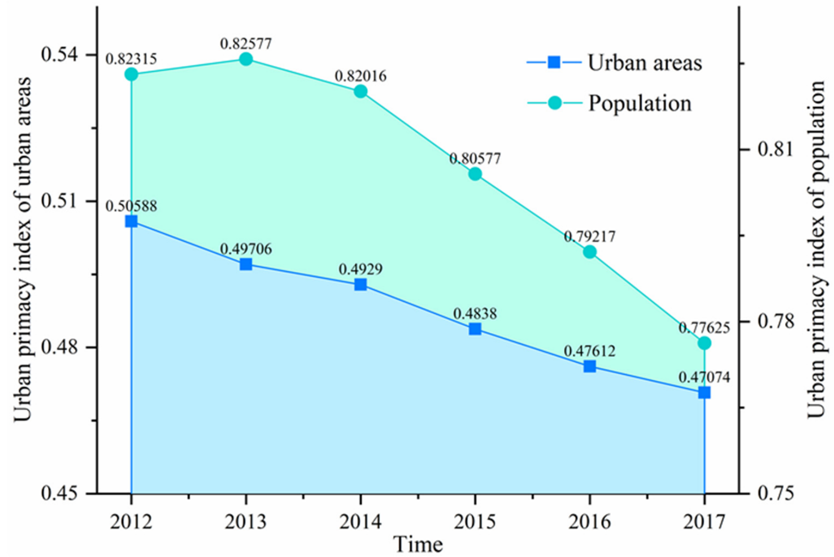


Figure 7. Variations in urban primacy index for extracted urban areas and the statistic population data of the YRD from 2012 to 2017.

3.2.3. Variations in the Gini Coefficient

In this study, the Gini coefficient for cities was used to analyze the degree of balance in city development in the study area. Figure 8 shows the variations in the Gini coefficient computed from the extracted urban areas using NTL remote-sensing and population data from 2012 to 2017. The value of the Gini coefficient calculated from the extracted urban areas ranged from 0.2 to 0.3, which indicates that the development of the city size in the study area occurred at a relatively balanced level during this period. Furthermore, the Gini coefficient displayed a gradually decreasing tendency from 2012 to 2017, with the value declining from 0.226 in 2012 to 0.215 in 2017. Such a decrease indicates that the city size of the YRD displayed a scattering trend from 2012 to 2017. Thus, the dispersed force in the study area was greater than the concentrated force, implying that the development of medium- and small-sized cities was faster than that of big cities, and the development of big cities reached saturation in the study area during this period.

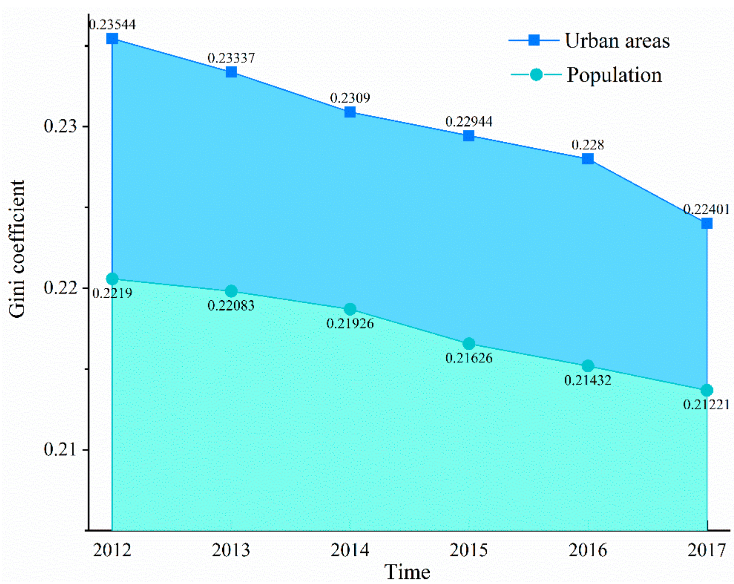


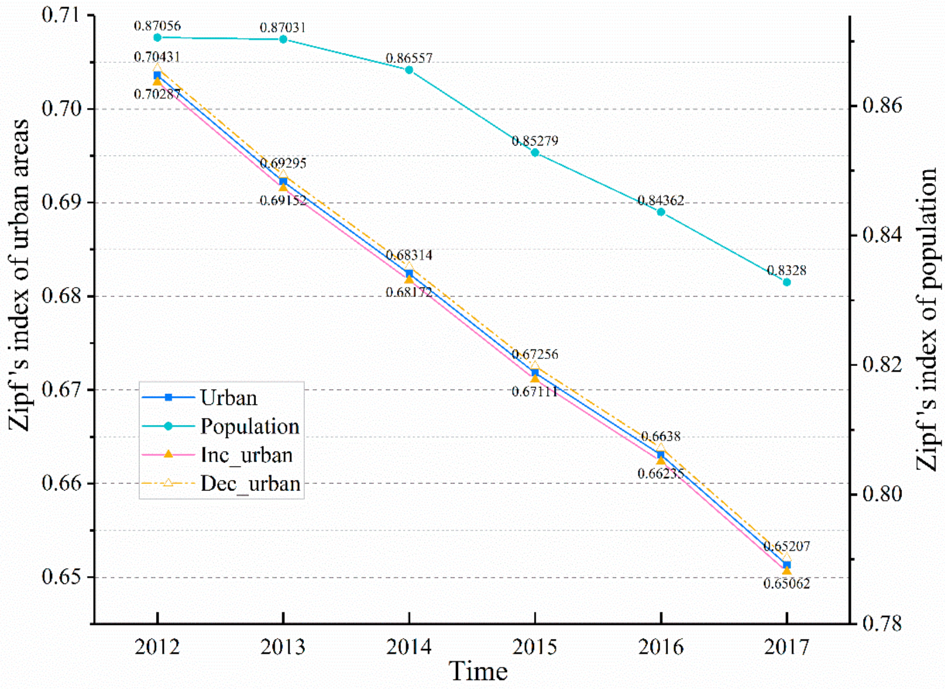
Figure 8. Variations in the Gini coefficient for extracted urban areas and the statistic population data of YRD from 2012 to 2017.

The Gini coefficient calculated from the population data was used to validate the value computed from the extracted urban areas of the study area. The value of the Gini coefficient calculated from population data also ranged from 0.2 to 0.3, and displayed a decreasing tendency from 2012 to 2017 (Figure 8), aligning with the variations in the Gini coefficient computed from the extracted urban areas. This result also highlights the feasibility of using extracted urban areas instead of population data to study the distribution and variation in city size.

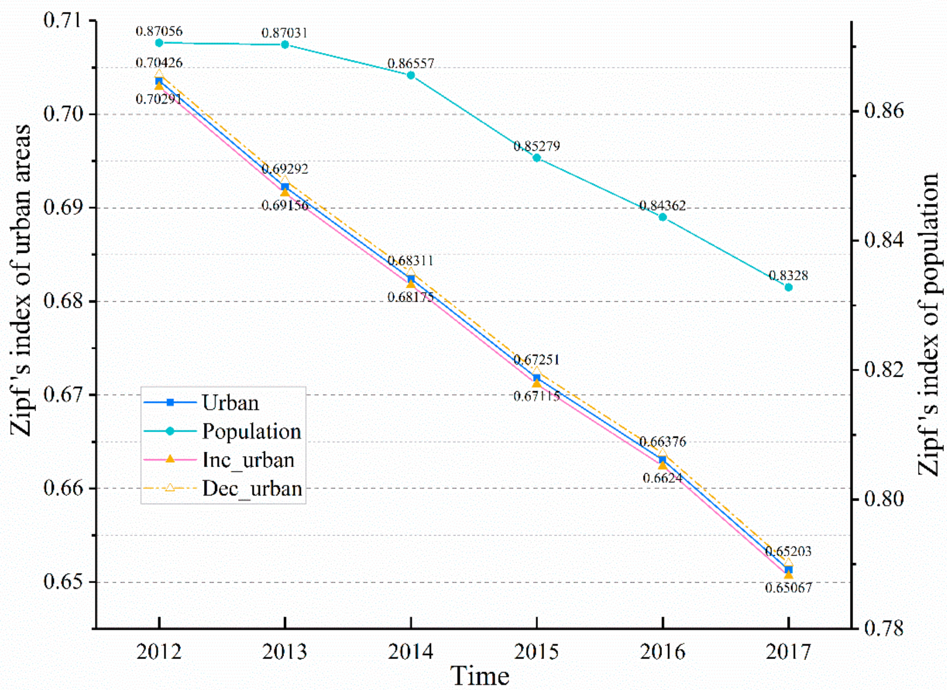
3.3. Sensitive Analysis

In the accuracy assessment of urban area extraction presented in Section 3.1.1, five small to medium-sized cities with lower levels of economic development, namely Chizhou, Xuancheng, Anqing, Chuzhou, and Yancheng, were found to have relatively low accuracy, with Kappa coefficients and geometric mean accuracy below 60% and 0.6, respectively. This finding was mainly due to cities with lower levels of economic development having relatively lower signals of NTL and LST images, which may affect the urban extraction results. However, the purpose of this study was to prove that urban areas extracted using remote-sensing technology can be utilized to analyze variations in city size instead of statistical population data. Whether the extraction accuracy can affect the results of the city size in the study area must be determined. Therefore, we adjusted the urban areas of the five cities by increasing or decreasing 5% of the extracted urban areas to explore the effect of extraction accuracy on city size monitoring. When the urban areas of one city were adjusted, those of the other four cities remained unchanged. Zipf's index, the urban primacy index, and the Gini coefficient were computed from the originally extracted urban areas; the urban areas increased by 5%, and urban areas decreased by 5%. The top 11 cities ranked by the acreage of urban areas remained unchanged regardless of whether they decreased or increased by 5%, which caused very slight changes in the Gini coefficient (between 0.0001–0.0003). Thus, only Zipf's index and the Gini coefficient were calculated (Figure 9a–e) and (Figure 10a–e).

As depicted in Figure 9a–e, the Zipf's index calculated from the urban areas increasing by 5% or decreasing by 5% was consistent with that computed from the original urban areas, with values below 1.0 from 2012 to 2017, presenting a gradually decreasing tendency. This finding indicates that the development of small and medium-sized cities was more prominent than that of big cities in the YRD, and the dispersed force was larger than the concentrated force in the YRD. Further, as depicted in Figure 10a–e, the Gini coefficient calculated from both urban areas increasing by 5% or decreasing by 5% was consistent with that computed from the original urban areas, with all values ranging from 0.2 to 0.3, highlighting a gradually decreasing tendency from 2012 to 2017. This finding indicates that the development of YRD has reached a relatively balanced level. Further, the city size of YRD tended to be scattered from 2012 to 2017. Regardless of a decrease or increase by 5% in the extracted urban areas, the difference between the two indices calculated from the original urban areas and the adjusted urban areas was very small, with values below 0.001 (the three lines of the two indices calculated from the original and adjusted urban areas were very close). Consequently, the relatively low accuracy of urban area extraction of some cities has little effect on the monitoring of variations in city size in this study.

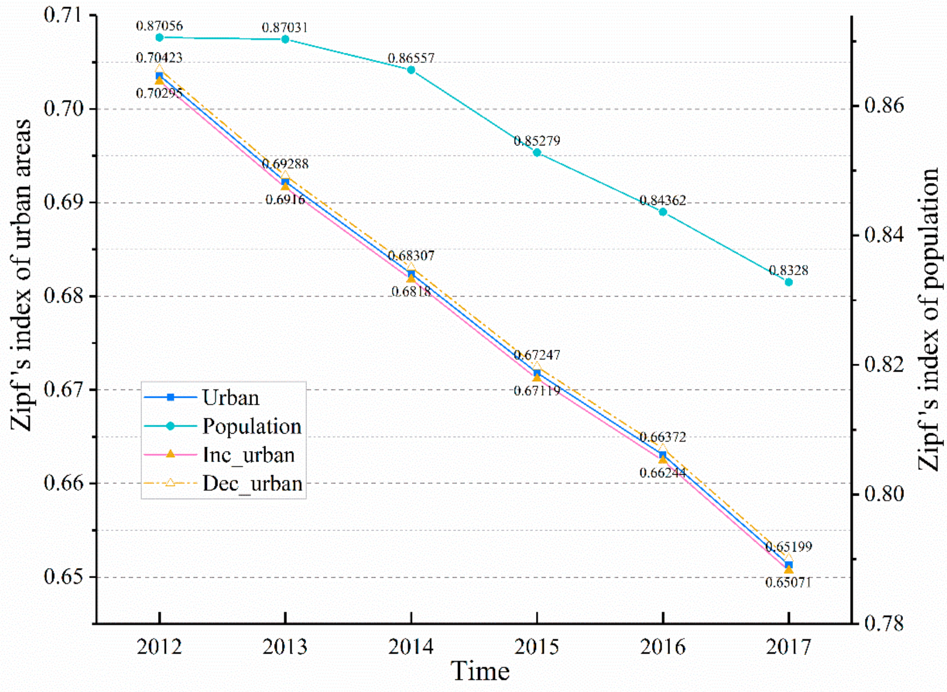


(a) Yancheng

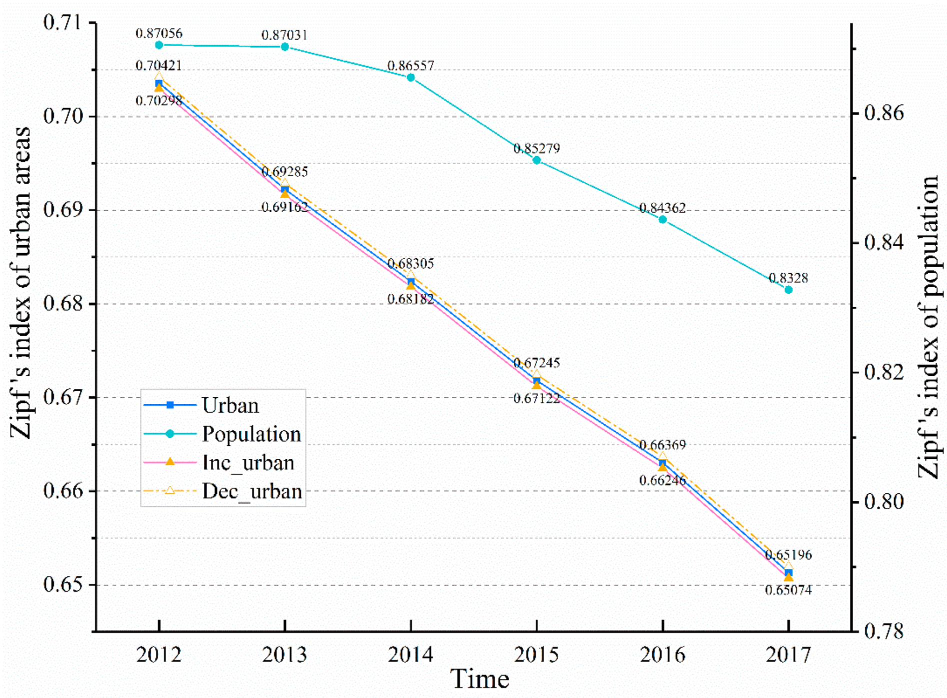


(b) Chuzhou

Figure 9. *Cont.*

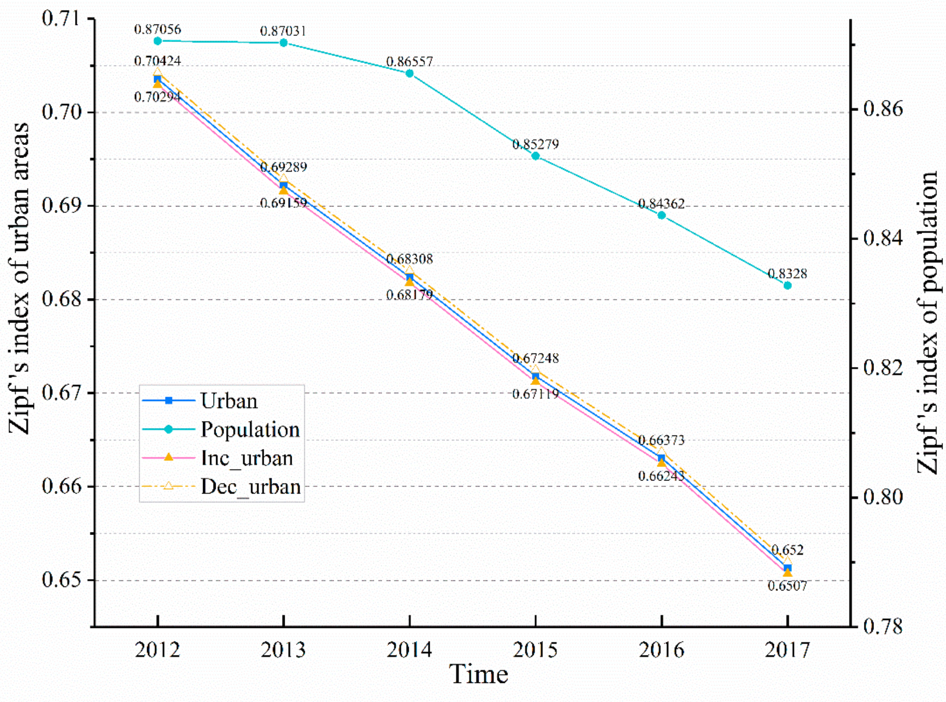


(c) Xuancheng



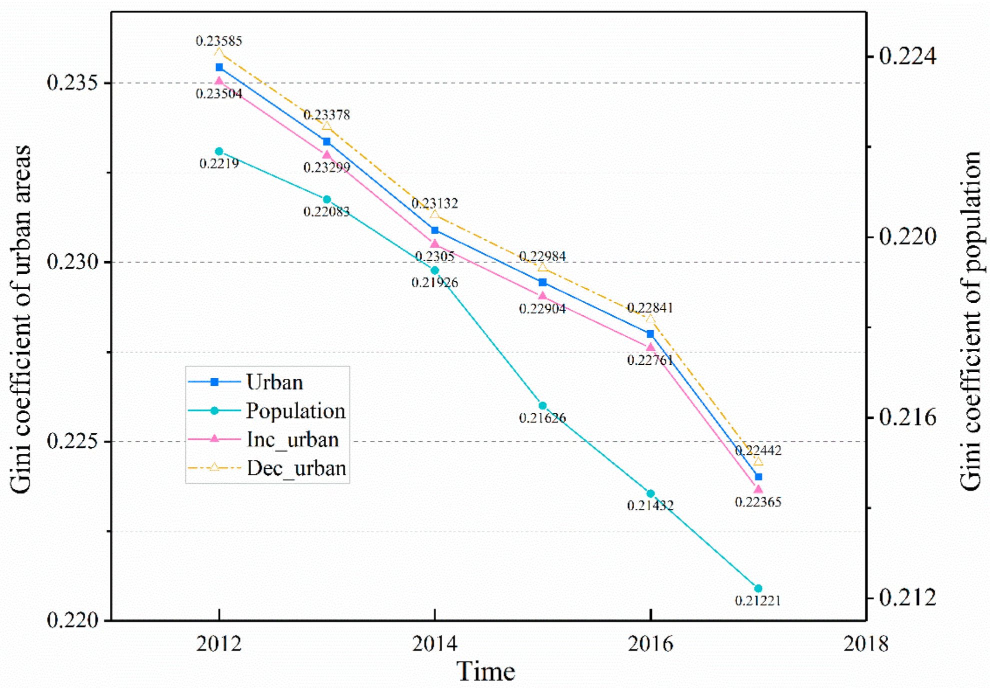
(d) Chizhou

Figure 9. *Cont.*



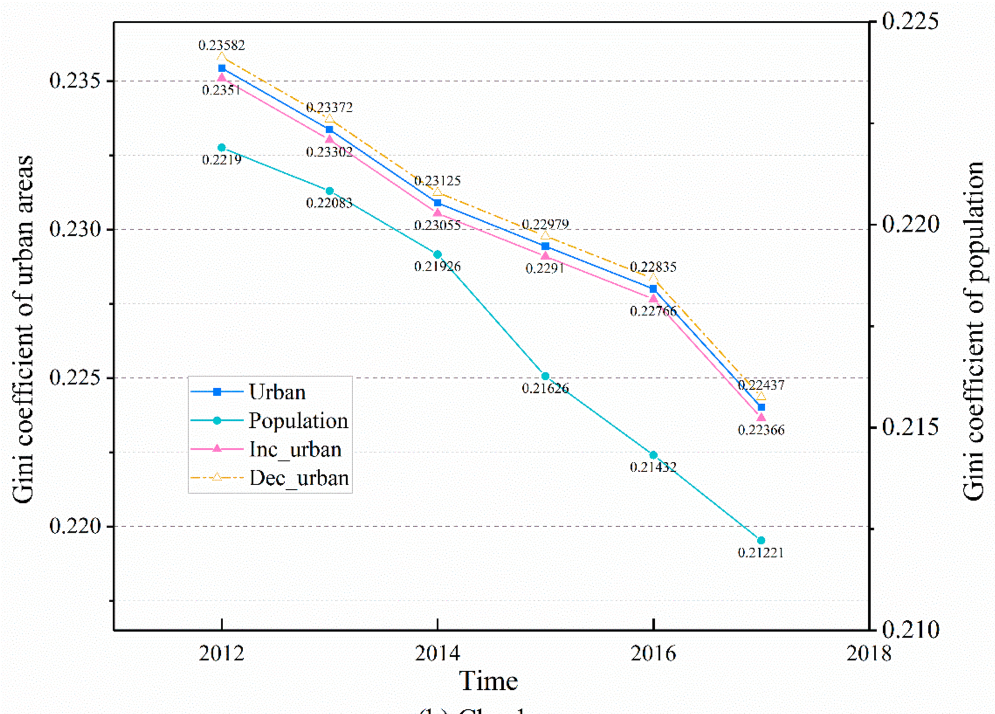
(e) Anqing

Figure 9. Variations in the Zipf's index of the YRD from 2012 to 2017; the "Inc_urban" and "Dec_urban" represent the results calculated from urban areas of five cities that increased by 5% and decreased by 5%, respectively.

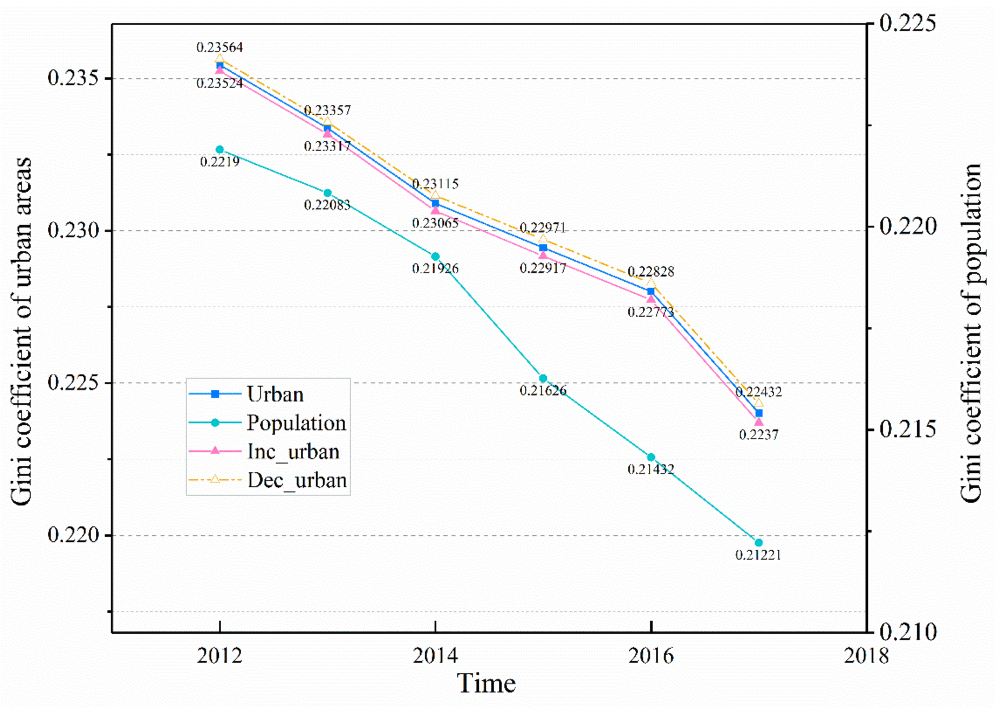


(a) Yancheng

Figure 10. *Cont.*

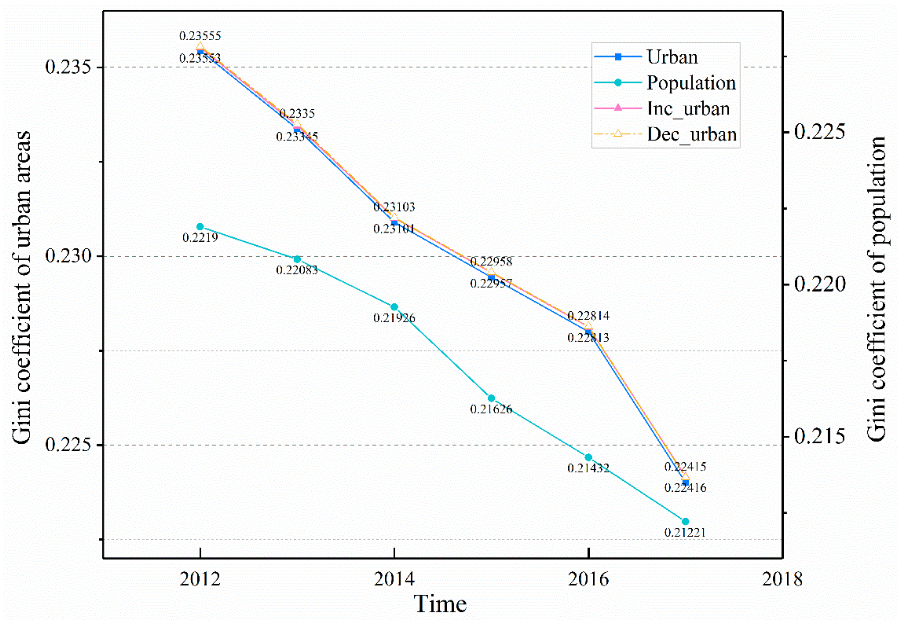


(b) Chuzhou

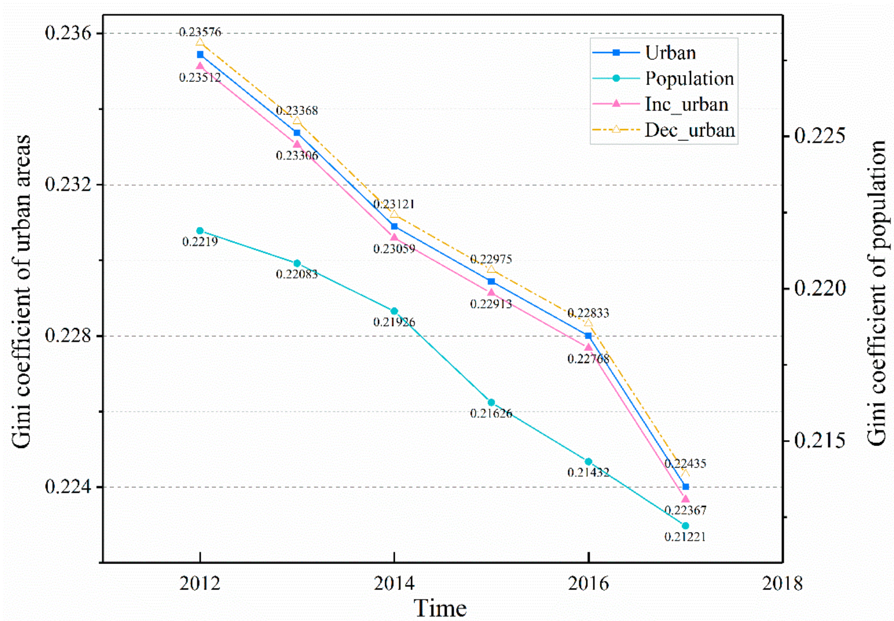


(c) Xuancheng

Figure 10. *Cont.*



(d) Chizhou



(e) Anqing

Figure 10. Variations in the Gini coefficient of the YRD from 2012 to 2017; The "Inc_urban" and "Dec_urban" represent the results calculated from urban areas of five cities that increased by 5% and decreased by 5%, respectively.

3.4. Comparison with Other Results

In this study, Zipf's law, the urban primacy index, and the Gini coefficient were computed from the statistical urban areas, and the distribution and variations in city size were compared with those calculated from the urban areas extracted from NTL remote sensing (Figures 11–13). As shown in Figure 11, Zipf's index calculated from statistical urban areas was consistent with that computed from the urban areas extracted from remote-sensing technology, with values ranging from 0.6 to 0.72 and a gradually decreasing tendency. Moreover, in Figure 12, the urban primacy index calculated from the statistic urban areas showed a decreasing trend, with values ranging from 0.41 to 0.51, thereby aligning with the values calculated from the extracted urban areas. The Gini coefficient calculated from statistical urban areas and extracted urban areas also showed some variation tendency in Figure 13, with values ranging from 0.2 to 0.3. Therefore, all three indices calculated from statistics and extracted urban areas revealed that the YRD did not fit the distribution of the primate city, the dispersed force was stronger than the concentrated force in the study area, and the development of the YRD reached a relatively balanced level. Consequently, the city size analyzed from the extracted urban areas had the same variation trend as that analyzed from the statistical urban areas, which proved the feasibility of utilizing NTL remote sensing to analyze the distribution and variation of city size instead of statistics.

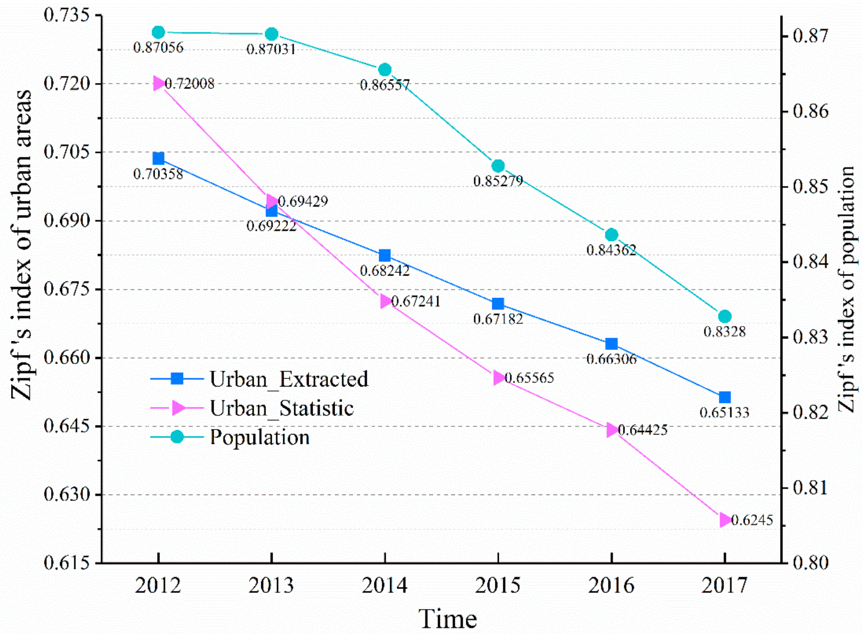


Figure 11. Variations in the Zipf's index for extracted urban areas, statistic urban areas, and population data of the YRD from 2012 to 2017.

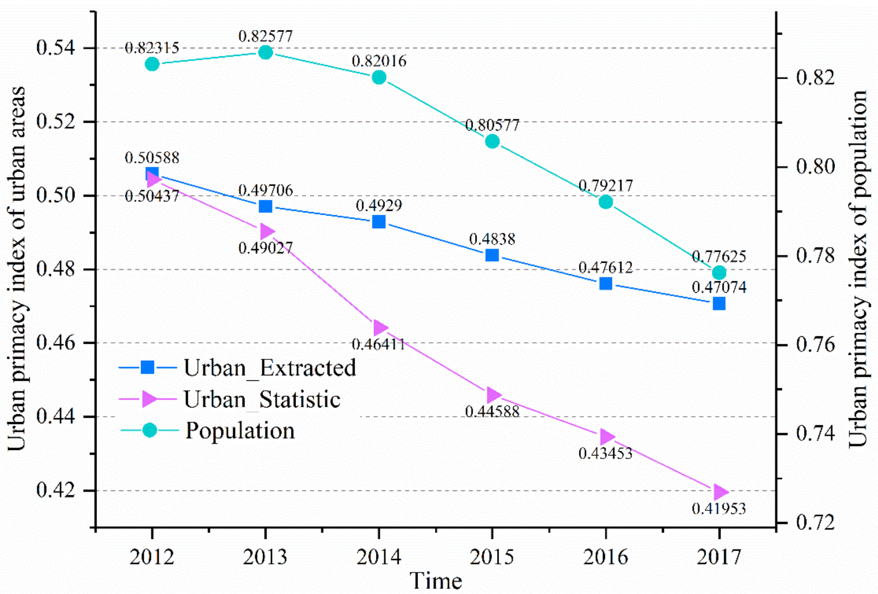


Figure 12. Variations in the urban primacy index for extracted urban areas, statistic urban areas, and population data of the YRD from 2012 to 2017.

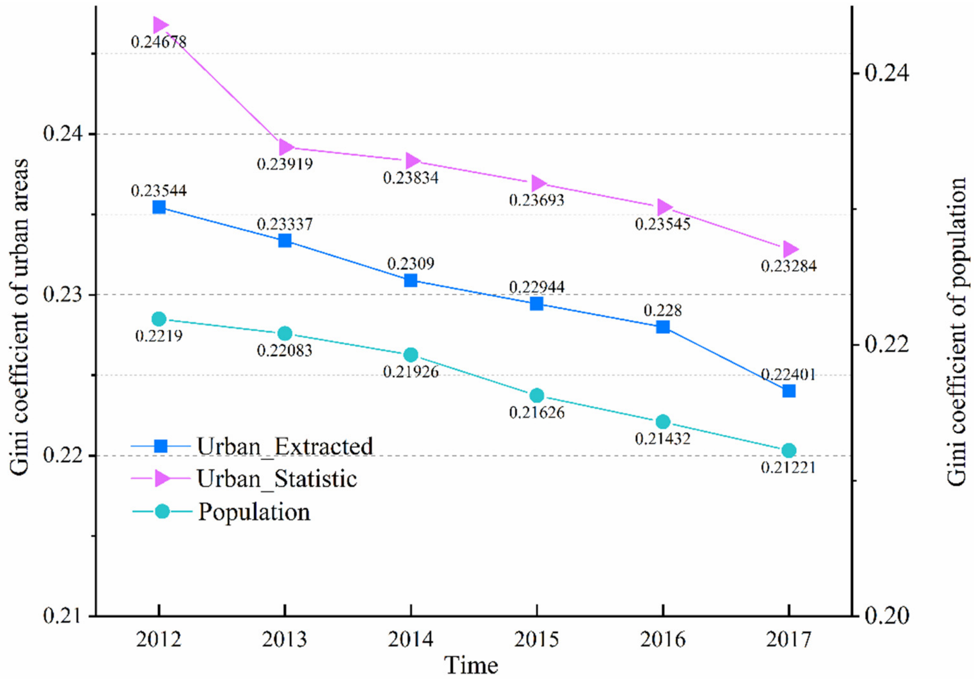


Figure 13. Variations in the Gini coefficient for extracted urban areas, statistic urban areas, and population data of the YRD from 2012 to 2017.

4. Conclusions

The YRD of China was selected as the study area for this assessment and multi-source remote-sensing data based on an ANN algorithm were employed to extract urban areas for the analysis of variations in city size from 2012 to 2017 in the study area, instead of the traditional approach of using population data. The main conclusions are as follows.

By combining NTL data with land-surface temperature and normalized differential vegetation index data, the method employed in this study could be used to efficiently extract the urban areas of the study area based on the ANN algorithm. The geometric mean accuracy and Kappa coefficient ranged from 79.21% to 84.88% and 0.761 to 0.828, respectively. In addition, the urban areas of the study area expanded gradually from 2012 to 2017, with the entire region and all cities in the region undergoing expansion. Moreover, the rank–size rule, urban primacy index, and Gini model were utilized to assess the development of city size in the study area from 2012 to 2017. Based on the results, the study area did not fit the distribution of the primate city, and the status and function of the primate city in the YRD gradually decreased during this period. The development of big cities was not prominent compared with that of small and medium-sized cities, and the dispersed force was greater than the concentrated force in the study area. The development of city size in the YRD urban agglomeration reached a relatively balanced level during this period. Additionally, the sensitivity analysis revealed that the relatively low extraction accuracy of urban areas for a few small cities had little effect on the results of city size in this study. The three indices calculated from the statistical population data and the statistical urban areas also displayed the same distribution and variation tendency as those computed from the extracted urban areas via NTL remote sensing. These findings indicate the applicability and feasibility of analyzing the variation in city size using NTL remote sensing.

Author Contributions: Conceptualization, Y.D. and J.H.; methodology, Y.D. and J.H.; investigation, Y.D., J.H. and W.M.; validation, Y.Y., S.J., Y.Z. and J.Z.; writing—original draft preparation, Y.D., J.H. and W.M.; writing—review and editing, Y.Y., S.J., X.P., Y.Z., K.C. and J.Z.; funding acquisition, Y.D. and Y.Y. All authors contributed to the analysis of the results and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Fundamental Research Funds for the Central Universities (Grant numbers B200202017, B210201013, and B220203008), the National Natural Science Foundation of China (Grant number 42071346), the Natural Science Foundation of Jiangsu Province under (Grant number BK20190495), and the Postgraduate Research and Practice Innovation Program of the Jiangsu Province (No. KYCX21_0529).

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

References

1. Schneider, A.; Friedl, M.A.; Potere, D. Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions'. *Remote Sens. Environ.* **2010**, *114*, 1733–1746. [CrossRef]
2. United Nations. World Urbanization Prospects: The 2014 Revisio. Available online: https://population.un.org/wup/publications/files/wup2014-report.pdf (accessed on 15 May 2020).
3. Muhammad, S.; Lean, H.H. Does Financial Development Increase Energy Consumption? The Role of In-dustrialization and Urbanization in Tunisia. *Energy Policy* **2012**, *40*, 473–479.
4. Walsh, C.J. Urban Impacts on the Ecology of Receiving Waters: A Framework for Assessment, Conservation and Restoration. *Hydrobiologia* **2000**, *2*, 107–114. [CrossRef]
5. Kalnay, E.; Cai, M. Impact of urbanization and land-use change on climate. *Nature* **2003**, *423*, 528–531. [CrossRef] [PubMed]
6. Weng, Q.; Firozjaei, M.K.; Sedighi, A.; Kiavarz, M.; Alavipanah, S.K. Statistical analysis of surface urban heat island intensity variations: A case study of Babol city, Iran. *GIScience Remote Sens.* **2018**, *56*, 576–604. [CrossRef]
7. Zipf, G.K. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*; Addison-Wesley: New York, NY, USA, 1949.
8. Mark, J. The Law of the Primate City. *Geogr. Rev.* **1939**, *2*, 226–232.
9. Gini, C. Measurement of Inequality of Incomes. *Econ. J.* **1921**, *31*, 124. [CrossRef]
10. Huang, Q.; He, C.; Gao, B.; Yang, Y.; Liu, Z.; Zhao, Y.; Dou, Y. Detecting the 20 year city-size dynamics in China with a rank clock approach and DMSP/OLS nighttime data. *Landsc. Urban Plan.* **2015**, *137*, 138–148. [CrossRef]
11. Anderson, G.; Ying, G. The Size Distribution of Chinese Cities. *Reg. Sci. Urban Econ.* **2005**, *6*, 756–776. [CrossRef]
12. Fan, C.C. The Vertical and Horizontal Expansions of China's City System. *Urban Geogr.* **1999**, *6*, 493–515. [CrossRef]
13. Xu, Z.; Zhu, N. City Size Distribution in China: Are Large Cities Dominant? *Urban Stud.* **2009**, *10*, 2159–2185. [CrossRef]
14. Levin, N.; Zhang, Q. A global analysis of factors controlling VIIRS nighttime light levels from densely populated areas. *Remote Sens. Environ.* **2017**, *190*, 366–382. [CrossRef]
15. Yao, Y.; Chen, D.; Chen, L.; Wang, H.; Guan, Q. A time series of urban extent in China using DSMP/OLS nighttime light data. *PLoS ONE* **2018**, *13*, e0198189. [CrossRef]
16. Zou, Y.; Peng, H.; Liu, G.; Yang, K.; Xie, Y.; Weng, Q. Monitoring Urban Clusters Expansion in the Middle Reaches of the Yangtze River, China, Using Time-Series Nighttime Light Images. *Remote Sens.* **2017**, *10*, 1007. [CrossRef]
17. Nitsch, V. Zipf Zipped. *J. Urban Econ.* **2005**, *1*, 86–100. [CrossRef]
18. Fu, Y.; Li, J.; Weng, Q.; Zheng, Q.; Li, L.; Dai, S.; Guo, B. Characterizing the spatial pattern of annual urban growth by using time series Landsat imagery. *Sci. Total Environ.* **2019**, *666*, 274–284. [CrossRef]
19. Huang, X.; Schneider, A.; Friedl, M.A. Mapping sub-pixel urban expansion in China using MODIS and DMSP/OLS nighttime lights. *Remote Sens. Environ.* **2016**, *175*, 92–108. [CrossRef]
20. Xie, Y.; Weng, Q.; Fu, P. Temporal Variations of Artificial Nighttime Lights and Their Implications for Ur-banization in the Conterminous United States, 2013–2017. *Remote Sens. Environ.* **2019**, *225*, 160–174. [CrossRef]
21. Yu, B.; Shu, S.; Liu, H.; Song, W.; Wu, J.; Wang, L.; Chen, Z. Object-based spatial cluster analysis of urban landscape pattern using nighttime light satellite images: A case study of China. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 2328–2355. [CrossRef]
22. Su, Y.; Chen, X.; Wang, C.; Zhang, H.; Liao, J.; Ye, Y.; Wang, C. A new method for extracting built-up urban areas using DMSP-OLS nighttime stable lights: A case study in the Pearl River Delta, southern China. *GIScience Remote Sens.* **2015**, *52*, 218–238. [CrossRef]
23. Small, C.; Elvidge, C.D. Night on Earth: Mapping Decadal Changes of Anthropogenic Night Light in Asia. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *22*, 40–52. [CrossRef]
24. Small, C.; Elvidge, C.D.; Balk, D.; Montgomery, M. Spatial scaling of stable night lights. *Remote Sens. Environ.* **2011**, *115*, 269–280. [CrossRef]
25. Elvidge, C.D.; Baugh, K.E.; Zhizhin, M.; Hsu, F.-C. Why VIIRS data are superior to DMSP for mapping nighttime lights. *Proc. Asia-Pac. Adv. Netw.* **2013**, *35*, 62. [CrossRef]
26. Miller, S.D.; Straka, W.; Mills, S.P.; Elvidge, C.D.; Lee, T.F.; Solbrig, J.; Walther, A.; Heidinger, A.K.; Weiss, S.C. Illuminating the Capabilities of the Suomi National Polar-Orbiting Partnership (NPP) Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band. *Remote Sens.* **2013**, *5*, 6717–6766. [CrossRef]
27. Cao, C.; Bai, Y. Quantitative Analysis of VIIRS DNB Nightlight Point Source for Light Power Estimation and Stability Monitoring. *Remote Sens.* **2014**, *6*, 11915–11935. [CrossRef]
28. Zhou, Y.; Smith, S.J.; Elvidge, C.D.; Zhao, K.; Thomson, A.; Imhoff, M. A cluster-based method to map urban area from DMSP/OLS nightlights. *Remote Sens. Environ.* **2014**, *147*, 173–185. [CrossRef]

29. Xie, Y.; Weng, Q. Updating Urban Extents with Nighttime Light Imagery by Using an Object-Based Thresh-olding Method. *Remote Sens. Environ.* **2016**, *187*, 1–13. [CrossRef]
30. Henderson, M.; Yeh, E.; Gong, P.; Elvidge, C.; Baugh, K. Validation of urban boundaries derived from global night-time satellite imagery. *Int. J. Remote Sens.* **2003**, *24*, 595–609. [CrossRef]
31. Liu, Z.; He, C.; Zhang, Q.; Huang, Q.; Yang, Y. Extracting the dynamics of urban expansion in China using DMSP-OLS nighttime light data from 1992 to 2008. *Landsc. Urban Plan.* **2012**, *106*, 62–72. [CrossRef]
32. Ma, T.; Zhou, Y.; Zhou, C.; Haynie, S.; Pei, T.; Xu, T. Night-Time Light Derived Estimation of Spa-tio-Temporal Characteristics of Urbanization Dynamics Using Dmsp/Ols Satellite Data. *Remote Sens. Environ.* **2015**, *158*, 453–464. [CrossRef]
33. Cao, X.; Chen, J.; Imura, H.; Higashi, O. A SVM-based method to extract urban areas from DMSP-OLS and SPOT VGT data. *Remote Sens. Environ.* **2009**, *113*, 2205–2209. [CrossRef]
34. Jing, W.; Yang, Y.; Yue, X.; Zhao, X. Mapping Urban Areas with Integration of DMSP/OLS Nighttime Light and MODIS Data Using Machine Learning Techniques. *Remote. Sens.* **2015**, *7*, 12419–12439. [CrossRef]
35. Xu, T.; Coco, G.; Gao, J. Extraction of urban built-up areas from nighttime lights using artificial neural network. *Geocarto Int.* **2019**, *35*, 1049–1066. [CrossRef]
36. Lu, D.; Tian, H.; Zhou, G.; Ge, H. Regional mapping of human settlements in southeastern China with multisensor remotely sensed data. *Remote Sens. Environ.* **2008**, *112*, 3668–3679. [CrossRef]
37. Yang, Y.; He, C.; Zhang, Q.; Han, L.; Du, S. Timely and accurate national-scale mapping of urban land in China using Defense Meteorological Satellite Program's Operational Linescan System nighttime stable light data. *J. Appl. Remote Sens.* **2013**, *7*, 073535. [CrossRef]
38. Kubat, M.; Holte, R.C.; Matwin, S. Machine Learning for the Detection of Oil Spills in Satellite Radar Images. *Mach. Learn.* **1998**, *30*, 195–215. [CrossRef]