An improved nightlight threshold method for revealing the spatiotemporal dynamics and driving forces of urban expansion in China

Cheng Huang \textsuperscript{a,b}, Qianlai Zhuang \textsuperscript{b}, Xing Meng \textsuperscript{a,c}, Hongwei Guo \textsuperscript{d}, Ji Han \textsuperscript{d,e,}\textsuperscript{*}

\textsuperscript{a} School of Geographical Sciences, East China Normal University, Shanghai, 200241, China
\textsuperscript{b} Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN, 47907, USA
\textsuperscript{c} Institute of Belt and Road & Global Development, East China Normal University, China
\textsuperscript{d} Shanghai Key Laboratory for Urban Ecological Processes and Eco-Restoration, School of Ecological and Environmental Sciences, East China Normal University, Shanghai, 200041, China
\textsuperscript{e} Institute of Eco-Chongming, 3663 N. Zhongshan Rd., Shanghai, 200062, China

ARTICLE INFO

Keywords:
Nighttime light
Dynamic threshold
Urban extraction
Driving forces
China

ABSTRACT

An accurate and efficient extraction of urban extent is important for understanding the dynamics of urban expansion process and for sustainable planning and management of cities. We proposed an improved dynamic nightlight threshold method to model urban extent and to reveal the spatiotemporal dynamics and driving forces of urban expansion. Differing from previous studies, we correct the blooming and over-saturation problems of nighttime light (NTL), and highlight a combination of NTL with urban population data for determining a yearly-continued and city-class-wide threshold for urban mapping. China is selected as a case study area to test the improved method and to gain insights to its urban expansion process. Through the validation, our method has been proven to be more accurate than the traditional NTL threshold method. Accordingly, the yearly-continued NTL data can better describe the changing patterns and driving forces of urban expansion than the yearly-discontinued land use and land cover data do. It is found that the total urban area in China has more than quadrupled from 25.2 thousand km\textsuperscript{2} in 1992 to 108.2 thousand km\textsuperscript{2} in 2013. Some significant pulses of urban expansion have been detected in our study, which may be attributed to the policy and socioeconomic impacts. Moreover, the panel regression based on annual NTL data indicates that GDP is a more important driver of urban expansion than urban population.

1. Introduction

Urban expansion has been recognized having substantial impacts on environment. It imposes environmental changes through the direct fragmentation of local landscape and the indirect change of biophysical properties of landscape that cause profound impacts on environment on different scales (Alberti and Marzluff, 2004; Han et al., 2017; He et al., 2017; Zhao et al., 2018; Liu et al., 2018a). Currently, more than 54% of the total world population (about 4 billion) lives in urban areas and that scale is supposed to continuously increase to 67% in 2050 (The world bank, 2016). Thus, an accurate and efficient modeling of the urban expansion processes and its spatial pattern has become increasingly important for understanding the urbanization-derived impacts, so that proper policy implications could be developed to forward both the local and global sustainability.

China has undergone a rapid and massive urbanization process. In the past decades, its urbanization rate measured by the percentage of the population living in urban area has more than tripled, from 18% in 1978 to 59% in 2017 (NBS, 2018). A great endeavor has been made on how to accurately and efficiently capture the urban expansion process since it is the premise of urban planning and sustainable city management. As one of the commonly used indicators, urban built-up area is recognized as the closest and most available measurement of urban socioeconomic activity and its geographical changes (Cai and Zhang, 2014; Wang et al., 2017). Currently, urban built-up data are generally from two sources. One is land use and land cover (LULC) data, which are usually classified from the day-time remote sensing images. The other is the modeled urban extent from the nighttime light (NTL) images.
The former one is recognized as the most accurate data that reflects the dynamic changes of urban extent both in time and in space. In regional and city-scaled studies, the medium and high-resolution satellite images such as Landsat TM/ETM+, IKONOS and Quickbird data are widely used for LULC analysis. Though the pre-processing of the raw data and the classification of the land types are usually time-consuming and require abundant professional knowledge and techniques, the development of cloud-computing platform, the improvement of machine learning techniques, and increasing accessibility of land use and cover data in recent years have greatly enhanced the efficiency of LULC classification. However, it is still difficult to separate the bare background from the built-up land by only using the day-time remote sensing images as they are loosely linked to the human activities. Moreover, the LULC data are usually provided with a specific time interval (for example, global LULC data provided by the National Geomatics Center of China (NGCC, 2015) are available with five-year interval), and not updated timely (Xie and Weng, 2016; Roy et al., 2014; Julien et al., 2011). Due to the significant changes occurred in China’s urbanization both in time and space, a yearly-continued and reliable spatial urban land data rather than a yearly-discontinued one would be more helpful to reveal the spatiotemporal dynamics of urban expansion.

The latter one uses NTL as a proxy for urban extent modeling. Urban boundary is extracted based on a threshold of NTL intensity. Though the accuracy of the modeled urban extent is usually lower than that from the day-time remote sensing image classification. The most advantage of using NTL for urban modeling is because of its close relationship to socioeconomic variables, such as GDP, population density, energy consumption. Thus, it is capable to remove the bare background of urban built-up land and reflect the urban land change driven by human activities. Moreover, since NTL data is provided in a long-term and yearly-continued form, it is important for understanding the urban expansion in a rapidly urbanizing region like China. Among all the NTL data sources, Defense Meteorological Satellite Program’s Operational Linescan System (DMSP-OLS) nighttime light data and the relevant methods for urban extraction have gained increasing attention and been widely used in urbanization studies (Zhou et al., 2014; Henderson et al., 2003; Huang et al., 2014). The DMSP-OLS NTL images are the nighttime lights detected from those shining earth’s surfaces such as buildings, roads, and parking lots (Elvidge et al., 1997). NASA released the yearly DMSP-OLS NTL data from 1992 to 2013, which enables the long-term monitoring and evaluation of urban expansion (Zhou et al., 2014). Threshold approach has also been proven as a simple and useful method for mapping urban area at regional, national and global scales (Ma et al., 2015; Amaral et al., 2005; Liu et al., 2012; Elvidge et al., 1999; Yi et al., 2014; Xie and Weng, 2016; Li and Chen, 2018; Meng et al., 2017).

Based on the NTL and threshold method, a great number of studies have been reported on China’s urban expansion issue. For example, Xu et al. (2016), Yi et al. (2014) and Ma et al. (2012) analyzed the dynamic change of China’s urban expansion on provincial and regional scales. Most of these studies assumed all cities had the same characteristics in NTL while neglecting the spatial and temporal variations of NTL threshold for urban mapping. Recently, there are several studies have paid attention to the heterogeneity issue of NTL threshold. For example, Zhou (2014) used clustering method to divide NTL into different city clusters, and then determined the threshold of each cluster. Xie (2016) proposed an object-based urban thresholding method for NTL to extract multi-temporal urban extent. However, due to the spillover and oversaturation problems of NTL data, it will inevitably cause the overestimation or underestimation of the modeled urban extent if the threshold is only determined by NTL data. In addition, though the temporal change of threshold has been considered in some studies, most of the existing threshold methods usually assign the estimated threshold of a specific year to those of adjacent years. In fact, the traditional threshold method can be used for mapping the yearly urban expansion. But the threshold itself is discontinuous dynamic rather than yearly-continued dynamic (Liu et al., 2012).

In sum, there are at least two aspects of the threshold method could to be improved substantially. First, the spatial variations of NTL.
thresholds for urban mapping should be taken into account by including not only the NTL intensity, economic zoning, but also the city size or even more socioeconomic features that may affect the regional difference. Second, the temporal variations of NTL thresholds can be further improved by integrating LULC data with socioeconomic data. Since urban expansion is driven by socioeconomic activities (e.g., GDP, population), the determination and extrapolation of thresholds should consider both the geophysical features and the socioeconomic conditions so as to generate an accurate and reliable result.

As such, considering the deficiencies of the existing studies, we proposed an improved dynamic threshold method to extract urban extent effectively based on NTL and urban population data. Differing from the static or discontinuous dynamic traditional threshold method, our method provides a yearly-continued and city-class-wide threshold for urban extraction.

2. Data and methods

2.1. Study area and data description

China is selected as a case study area to test our proposed method, and to gain insights to its urban expansion. Both statistical and spatial data are collected for analysis. The statistical data include urban population, consumption level, Gross Domestic Product (GDP) of the secondary and tertiary industries, which are published by the China City Statistical Yearbook (NBS, 2014), and China Statistical Yearbook (NBS, 2014). Urban population refers to the permanent resident population, which has been residing in city for over half a year. All the GDP data were converted to the 1978 price in order to overcome the inflation over time. Totally 336 China cities were classified into five classes according to the urban classification standard released by the State Council of the People’s Republic of China (SCPRC, 2014). The criteria are as follows: 1) Small city: urban population less than 500 thousand; 2) Medium city: urban population between 500 thousand and 1 million; 3) Large city: urban population from 1 million to 5 million; 4) Mega city: urban population between 500 thousand and 1 million; and 5) Supermega city: urban population larger than 10 million. Fig. 1 shows the spatial distribution of the five city classes.

The spatial data included LULC, NTL, and administrative boundary map. The administrative boundary was collected from the National Geomatics Center of China. LULC of 1990–2010 with 5-year interval were obtained from the Data Center for Resources and Environmental Sciences and the Chinese Academy of Sciences (http://www.resdc.cn), which was processed via manual visual interpretation based on the Landsat TM/ETM + images (RESDC, 2016). Generally, LULC was classified into six categories, including cropland, forest, grass, water, urban, and unused. In this study, we use urban land as a basis for determining the NTL threshold. Yearly NTL data in 1992–2013 were obtained from the National Centers for Environment Information (NCEI, 2017) of National Oceanic and Atmospheric Administration.

2.2. Improved dynamic threshold method for urban mapping

The traditional threshold method, that is widely employed in a number of studies (for example, Gao et al., 2015; Liu et al., 2012; Lu et al., 2018) usually assigns the estimated threshold of a specific year to those of adjacent years. For example, Liu et al. (2012) deduced the threshold based on the LULC data for the year 2000. Then the thresholds of 1998, 1999, 2001, 2002 were assumed to be the same with that of 2000. Recently, temporally extrapolated thresholds per urban cluster was used to analysis the dynamic changes of urban extent (Xie and Weng, 2016, 2017). Differing from the traditional one, we proposed an improved method highlighting a combination of NTL with urban population or GDP data for determining a yearly-continued and city-class-wide threshold for urban extent extraction.

Fig.S1 in the Supplementary Material shows the flowchart of threshold determination. First, NTL data were calibrated to correct the blooming and over-saturation problems following the methods proposed by previous studies (Elvidge et al., 1997, 2009; Liu et al., 2012). After the pre-processing, all the NTL data become comparable and consistent, and the highest digital number (DN) value increased from 63 to 78.16. All the spatial data were projected into the Krasovsky_1940 projection. Second, urban area from LULC and NTL were used to calculate the largest Kappa index for each city to determine the threshold in 1995, 2000, 2005, and 2010 respectively. Then, the average DN value of each city class can be decided as threshold for urban extraction. Here, Kappa is a index for evaluating classification accuracy, as shown in Eq (1). The closer to 1, the higher the accuracy is (Viera and Garrett, 2005; Kier, 1987).

\[ K_i = \frac{(P_{i|0} - P_{i|1})}{(1 - P_{i|0})} \]  

where, \( i \) represents city, city class or the whole research area, which depends on the purpose of the validation. \( E \) is the number of pixels that are urban areas in both the NTL and LULC datasets. \( A \) is the total number of pixels. \( L_{(i,t)} \) and \( N_{(i,t)} \) are the number of the pixels of urban areas of \( i \) from LULC and NTL respectively. \( L_{(0)} \) and \( N_{(0)} \) are the number of the pixels of non-urban areas of \( i \) from LULC and NTL respectively. \( P_{(i,0)} \) and \( P_{(i,1)} \) are intermediate variables.

Third, since a great number of literatures have proven that urban population and GDP are two important driving forces of urbanization (Li and Li, 2019; Han et al., 2009; Tan et al., 2018; Xia et al., 2019), which eventually affects the urban boundary expansion and the NTL value of a city. The increase of the size and activity intensity of urban population will also cause the NTL value increase. Since the NTL threshold is determined by the urban land that is classified from the day-time remote sensing images. Thus, it can be assumed that urban population drives the change of NTL. The thresholds of 1995, 2000, 2005, and 2010 were determined based on LULC. The threshold for those years during 1992–1997, 1998–2002, 2003–2007, and 2008–2013 is modeled through the change rate of urban population or GDP. Since the thresholds at year \( t \) and \( t+5 \) are fixed. To keep the modeled threshold in \( t+5 \) match with that from LULC, an adjustment parameter is adopted, which is shown as follows.

\[ TS_{i,(t+5)} = TS_{LUC} \times \left(1 + \frac{R_{(i,t)}}{R_{(i,t+5)}}\right) \]  

where, \( TS \) is the modeled dynamic threshold value; \( TS_{LUC} \) is the threshold value determined based on the land use change data. The subscript \( t \) means year; \( R \) is the change rate of urban population or GDP.

The model results were validated through the following sequence. First, we downloaded the urban land data of China in 2010 from Liu et al. (2018b), and resampled the data resolution from 30m to 1 km by using the majority raster resampling method through ArcMap 10.0. Their result was estimated based on the Landsat images and Google Earth Engine Platform, which is different from the LULC used in our analysis. It can be acknowledged as an independent data source and used for validating our model’s accuracy. Second, the extracted urban extent was compared by using the traditional threshold method and by using our improved method. The overall accuracy (OA, as shown in Eq. (3)) and Kappa indices were calculated for both methods. The higher OA and Kappa is, the higher accuracy the method will have.

\[ OA = \frac{\sum C_i}{A} \]  

where, \( A \) is the pixel number of urban land, and \( C_i \) is the pixel number of urban \( i \) that was classified correctly.
2.3 Driving force analysis

Based on the derived urban area, a panel regression model, as shown in Eq. (4), was developed to investigate the driving forces of urban expansion. Considering the city-level data availability those explanatory variables used in literature (Chen et al., 2018; Wang et al., 2018), GDP, population, and consumption level are selected. In addition, the panel regression result based on our method was compared with that based on LULC, so as to test the reliability of our proposed method. The natural logarithms of both sides of the regression model are conducted to minimize the heteroscedasticity.

\[
\ln(\text{UA}) = \beta_1 \ln(\text{GDP}) + \beta_2 \ln(\text{UPOP}) + \beta_3 \ln(\text{CON}) + C
\]

where, \(\text{UA}\) represents urban area, GDP and UPOP represents gross domestic product and urban population. CON is consumption level, defined as the money consumed per year.

3. Results and discussions

3.1 Method validation and sensitivity analysis

As illustrated in Table 1, OA of our proposed improved dynamic threshold method is significantly higher than that of the traditional threshold method. The total OA of our method in each city class was close to 90%. Moreover, OA of urban population-based method was relatively higher than GDP-based one. Kappa index of our methods and the traditional method are close to each other and all higher than 0.3, which suggests our method and traditional method are valid for urban area extraction (Li et al., 2020). The total Kappa of urban population-based method was relatively higher than GDP-based one. In sum, the improved threshold method is valid, and with relatively higher accuracy especially using the urban population for yearly-continued threshold modeling.

In addition to the validity and accuracy test, we performed the sensitivity analysis of both the threshold and urban area to the change of urban population and GDP. As shown in Table S1 in the supplementary material, the threshold and urban area are both of relatively low sensitivity to the change of urban Population and GDP growth rate. The relatively lower sensitivity of urban area to the change rate of urban population indicated that the modeling method based on urban population may be more robust than the method based on GDP.

3.2 Spatiotemporal patterns of urban expansion

Fig. 2 demonstrates the long-term changing pattern of China’s urban expansion. In Fig. 2a, we compared the urban area from our method based on urban population (TS_POP) and from the traditional threshold method (TS_Traditional); (b) Urban area by city class from TS_POP.

Table 1
Accuracy test of the traditional threshold method and improved dynamic threshold method.

<table>
<thead>
<tr>
<th>Method</th>
<th>Supermega city</th>
<th>Mega city</th>
<th>Large city</th>
<th>Medium city</th>
<th>Small city</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>OA</td>
<td>TS_Traditional</td>
<td>84.15%</td>
<td>69.53%</td>
<td>63.97%</td>
<td>52.94%</td>
<td>47.59%</td>
</tr>
<tr>
<td></td>
<td>TS_POP</td>
<td>90.89%</td>
<td>86.34%</td>
<td>91.08%</td>
<td>90.83%</td>
<td>87.44%</td>
</tr>
<tr>
<td></td>
<td>TS_GDP</td>
<td>90.49%</td>
<td>86.38%</td>
<td>90.92%</td>
<td>89.77%</td>
<td>84.41%</td>
</tr>
<tr>
<td>Kappa</td>
<td>TS_Traditional</td>
<td>0.57</td>
<td>0.48</td>
<td>0.44</td>
<td>0.36</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>TS_POP</td>
<td>0.62</td>
<td>0.51</td>
<td>0.47</td>
<td>0.44</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>TS_GDP</td>
<td>0.55</td>
<td>0.45</td>
<td>0.50</td>
<td>0.41</td>
<td>0.30</td>
</tr>
</tbody>
</table>

TS_Traditional: Traditional threshold method. It assigns the estimated threshold of a specific year to those of adjacent years (Gao et al., 2015; Liu et al., 2012; Lu et al., 2018).

TS_POP: Improved dynamic threshold method based on urban population.

TS_GDP: Improved dynamic threshold method based on urban population.
land policies (Zhou et al., 2013; Wei et al., 2017). To make sure the significant pulse was consistent with the real pattern of China’s urbanization rather than causing by the temporal inconsistency of NTL data, we randomly picked up 5 cities of each city class, and calculated the growth rate of urban area based on the improved, traditional threshold methods, and based on LULC data. As shown in Table S2 in the supplementary material, the growth rate of our method was larger than that of LULC. While the growth rate of traditional threshold method was much smaller than that of Landsat. The average growth rate of all the randomly picked cities based on our method was more consistent with that from Landsat data, which suggests the improved threshold method can better describe the dynamic pattern of China’s urbanization than the traditional threshold method does.

When looking at the dynamics of the urban expansion by city class, Fig. 2b) indicates that the total urban area in China has been growing at an unprecedented speed of 7.19% per annum, from 25.16 thousand km\(^2\) in 1992 to 108.18 thousand km\(^2\) in 2013. Among them, large cities took the largest proportion (40.43% in 2013), followed by the Medium cities.

![Spatial distribution of urban areas in China](image)
The small and medium cities in the inland regions grew at an unprecedented speed in the past decades. Unbalanced economic development is one of the factors of regional differences in urban sprawl in China. After the reform and opening-up, foreign investment entered the eastern region, which promoted the economic prosperity of the eastern. Moreover, land and tax policies have stimulated a boom in land markets and accelerated urban sprawl since 1990, especially the rapid development of small and medium-sized cities in the west. (Wei et al., 2017; Huang et al., 2015, 2016). Although the inequality of urban development still exists, the gap is gradually narrowing after the extensive development of China’s western region. Fig. 3 shows the spatial distribution and variation of urban areas in China. Generally, an uneven spatial distribution of urban areas can be detected (Fig. 3a). About 80% of the supermega cities, 29% of mega cities, and 36% of large cities were located in seven provinces and three municipalities in the eastern and coastal part of China. It is to large extent affected by the economic development. Since the China’s reform and opening-up started earlier in the eastern and coastal regions than in the inland regions. Thus, the urbanization occurred more significantly in these economically wealthy areas. Fig. 3b illustrates the average annual growth rate urban area in each city from 1992 to 2013, and classified the change into six grades according to the natural breaking methods (Brewer and Pickle, 2002). It is found that the fast growth occurred in the Quzhou, followed by Suqian and Taizhou with their annual growth rate more than 20%. Owing to the natural endowment, the well development of tourism industry in Quzhou contributes a lot to the urban sprawl as it attracts great number of people to visit and even live here (Li et al., 2019). For the Suqian city, the construction industry plays an important role in the annual GDP, which results in the rapid expansion of urban area (Zhang et al., 2014). Industrial output accounts for 38.75% of GDP, of which construction accounts for 8.35%. While the Taizhou city, it is an import harbor in the eastern China. the large-scaled trading stimulates the economic development and the urbanization (Qian and Shi, 2008).

3.3. Driving forces of urban expansion

This study adopts panel analysis method to explore the driving force of urban expansion. Augmented Dickey–Fuller (ADF) and Phillips–Perron (PP) methods are used for unit root test. The result of unit root test (Table S3 in the supplementary material) shows that the time series data is stable, so the regression model can be established.

According to the Hausman test, the fixed-effect model was selected. Feasible generalized least-squares (Feasible GLS) method was adopted to estimate the static panel model. Table 2 shows the results of panel data regression analysis using LULC five-year interval data and annual NTL data. It is found that the regression based on the two different dataset both have a relatively high R², which means the panel regression model has an explanatory power and the results are reliable. However, the panel regression based on the annual NTL data showed that GDP is the most important driver of urban expansion, followed by population. It is consistent with several existing studies, such as Chen et al. (2018), Abiodun et al. (2017). While the regression based on the five-year interval data suggested that population is the most important driving factor, followed by GDP.

4. Conclusions

In this study, we propose an improved dynamic NTL threshold method for extracting the urban extent, and revealing the spatiotemporal patterns and driving forces of urban expansion. China is selected as a case study area for testing the method and gaining insights to the urbanization process. Differing from the previous studies, we highlight a combination of NTL with urban population data for determining the yearly-continued and city-class-wide thresholds for urban mapping. Through the validation, our method is proven to be more accurate than the traditional NTL threshold method in extracting the annual urban extent for most of the city classes. The overall accuracy of our method reached 71.7% while it was 62.3% for the traditional threshold method. Our method could be a supplement to the existing NTL-based approaches for modeling the urban land expansion. The major findings of China’s case study using the improved method are as follows:

First, the total urban area of China has increased by $3.8 \times 10^3$ km² from 1992 to 2013 with an annual growth rate reached 7.2%. Some significant pulses of urban expansion especially in 2003 have been detected by the improved threshold method. It was proven that our method can better reveal the pattern of urban expansion than the traditional threshold method does.

Second, the speed of urban expansion varied from different city sizes. The small cities grew faster than the other city classes with an average annual rate of 8.4%. The inclined policies, such as the national five-year plans, and develop-the-west plan, give preferences and financial support to those small and medium cities making their growth speed faster than the large and megacities.

Third, the panel data regression based on annual NTL data indicates that GDP is a more important driver of urban expansion than urban population, which is consistent with the existing studies that also study the driving forces of China’s urbanization. While the regression based on the five-year interval LULC data suggests a reversed result. It again indicates that a yearly-continued NTL data rather than a yearly-discontinued LULC data would be more helpful to reveal the driving forces of urban expansion in a rapidly urbanizing area like China.

Though this research has the merit in modeling urban extent, it still has some limitations that can be improved in the future works. The threshold for extracting urban was available according the availability of LULC data in limited years (for example, 1995, 2000, 2005, and 2010 in this study). While the threshold for rest years was estimated value. More LULC data from the daytime remote sensing are required to improve the overall accuracy of urban mapping. Both the spatial variation and temporal changes in NTL threshold should be considered in the future study in order to further improve the model accuracy.

Credit statement

Cheng Huang: Conceptualization, Methodology, Qianlai Zhuang:
Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Thanks to anonymous reviewers and editors for their valuable suggestions. This research was supported by the National Key R&D Program of China [2017YFC0505703]; National Natural Science Foundation of China [41401638, 41801314]; Shanghai Committee of Science and Technology Fund [19DZ1203503]; Shanghai Philosophy and Social Sciences Planning Project [2020BC009]; Shanghai Key Lab for Urban Ecological Processes and Eco-Restoration and the Fundamental Research Funds for the Central Universities [SHUES2020C01]; and the Institute for Belt and Road & Global Development Project [ECNU-BRGD-201803]; Cheng Huang acknowledges financial support from China Scholarship Council [201706140147]; and logistics support from Purdue University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvman.2021.112574.

References


