ENVIRONMENTAL RESEARCH
LETTERS

LETTER • **OPEN ACCESS**

Mapping global urban boundaries from the global artificial impervious area (GAIA) data

To cite this article: Xuecao Li *et al* 2020 *Environ. Res. Lett.* **15** 094044

Environmental Research Letters

LETTER

Mapping global urban boundaries from the global artificial impervious area (GAIA) data

Xuecao Li[1], Peng Gong[2,3,4,5], Yuyu Zhou[1], Jie Wang[6,7], Yuqi Bai[2,3,4], Bin Chen[8], Tengyun Hu[9], Yixiong Xiao[4], Bing Xu[2,3,4], Jun Yang[2,3,4], Xiaoping Liu[10], Wenjia Cai[2,3,4], Huabing Huang[11], Tinghai Wu[12], Xi Wang[7], Peng Lin[4], Xun Li[10], Jin Chen[13], Chunyang He[14], Xia Li[15], Le Yu[2,3,4], Nicholas Clinton[16] and Zhiliang Zhu[17]

[1] Department of Geological and Atmospheric Sciences, Iowa State University, Ames, IA 50011, United States of America
[2] Ministry of Education Key Laboratory of Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, People's Republic of China
[3] Center for Healthy Cities, Institute for China Sustainable Urbanization, Tsinghua University, Beijing 100084, People's Republic of China
[4] Tsinghua Urban Institute, Tsinghua University, Beijing 100084, People's Republic of China
[5] Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720-3110, United States of America
[6] State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, People's Republic of China
[7] AI for Earth Lab, Cross-Strait Institute, Tsinghua University, Beijing 100084, People's Republic of China
[8] Department of Land, Air and Water Resources, University of California, Davis, CA 95616-8627, United States of America
[9] Beijing Municipal Institute of City Planning and Design, Beijing 100045, People's Republic of China
[10] School of Geography and Planning, Sun Yat-Sen University, Guangzhou 510275, People's Republic of China
[11] School of Geospatial Engineering and Science, Sun Yat-Sen University, Guangzhou 510275, People's Republic of China
[12] School of Architecture, Tsinghua University, Beijing 100084, People's Republic of China
[13] State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, People's Republic of China
[14] Center for Human-Environment System Sustainability, State Key Laboratory of Earth Surface Process and Resource Ecology, Beijing Normal University, Beijing 100875, People's Republic of China
[15] School of Geographic Sciences, and Key Lab of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai 200241, People's Republic of China
[16] Google LLC, Mountain View, CA 94043, United States of America
[17] US Geological Survey, Reston, VA 20192, United States of America

E-mail: penggong@tsinghua.edu.cn **and** yuyuzhou@iastate.edu

OPEN ACCESS

RECEIVED
17 January 2020

REVISED
2 June 2020

ACCEPTED FOR PUBLICATION
11 June 2020

PUBLISHED
24 August 2020

Keywords: cellular automata, nighttime light, kernel density, multi-temporal, urban clusters, GEE

Supplementary material for this article is available online

Abstract

Urban boundaries, an essential property of cities, are widely used in many urban studies. However, extracting urban boundaries from satellite images is still a great challenge, especially at a global scale and a fine resolution. In this study, we developed an automatic delineation framework to generate a multi-temporal dataset of global urban boundaries (GUB) using 30 m global artificial impervious area (GAIA) data. First, we delineated an initial urban boundary by filling inner non-urban areas of each city. A kernel density estimation approach and cellular-automata based urban growth modeling were jointly used in this step. Second, we improved the initial urban boundaries around urban fringe areas, using a morphological approach by dilating and eroding the derived urban extent. We implemented this delineation on the Google Earth Engine platform and generated a 30 m resolution global urban boundary dataset in seven representative years (i.e. 1990, 1995, 2000, 2005, 2010, 2015, and 2018). Our extracted urban boundaries show a good agreement with results derived from nighttime light data and human interpretation, and they can well delineate the urban extent of cities when compared with high-resolution Google Earth images. The total area of 65 582 GUBs, each of which exceeds 1 km^2, is 809 664 km^2 in 2018. The impervious surface areas account for approximately 60% of the total. From 1990 to 2018, the proportion of impervious areas in delineated boundaries increased from 53% to 60%, suggesting a compact urban growth over the past decades. We found that the United States has the highest per capita

urban area (i.e. more than 900 m^2) among the top 10 most urbanized nations in 2018. This dataset provides a physical boundary of urban areas that can be used to study the impact of urbanization on food security, biodiversity, climate change, and urban health. The GUB dataset can be accessed from http://data.ess.tsinghua.edu.cn.

1. Introduction

Our planet is experiencing unprecedented global urbanization. The current rate of more than 50% of the world population living in urban areas is projected to reach ~70% by the middle of this century (United Nations 2019). One of the direct consequences of urban population growth is the rapid expansion of urban areas, which causes large tracts of cropland loss and deforestation (Foley *et al* 2005, 2011, DeFries *et al* 2010, Xi *et al* 2016), and creates great barriers to achieving the sustainable development goals (SDGs), particularly in developing regions such as Asia and Africa (Mbow *et al* 2017). Also, the rapid urban expansion has a great impact on urban heat island (Zhou *et al* 2004, Clinton and Gong 2013), building energy use (Güneralp *et al* 2017), flooding (Li *et al* 2016, Zhang *et al* 2018), air pollution (Gong *et al* 2012, Zhang *et al* 2012), and urban ecosystem (Alberti *et al* 2017, Li *et al* 2017, Wang *et al* 2019).

Accurate information on the dynamics of the global urban extent is highly needed for sustainable development goals (Lu *et al* 2015, Li and Gong 2016b, Li *et al* 2019). For example, predicting future urban development is a premise to evaluate the potential urbanization impact on our environment (Li *et al* 2019). Therefore, spatially explicit delineations of urban boundaries are crucial for a better understanding of the global urbanization.

Although it is possible to delineate urban boundaries through field surveys, it is not feasible to consistently carry out such activities in different cities over large areas. Satellite data have become a primary source for monitoring urban dynamics, with continuous observations spanning over the years to decades. Representative urban extent data have been derived from the nighttime light (NTL) data from the Defense Meteorological Satellite Program (DMSP) (Elvidge *et al* 2007, Zhou *et al* 2018), optical satellite observations from the moderate resolution imaging spectroradiometer (MODIS) (Schneider *et al* 2010), Landsat (Gong *et al* 2013, 2019a, Pesaresi *et al* 2015, Li *et al* 2018, 2020b, Liu *et al* 2018), Sentinel (Gong *et al* 2019b), and synthetic aperture radar data (Esch *et al* 2013, Li *et al* 2020a). Due to the difference in launch time and detection approaches, the derived urban extent maps from data on board of these various satellites are different in terms of their definitions, dates, and regions (Liu *et al* 2014), which limit their application in global change studies. Recently, a new dataset named global annual impervious area (GAIA) was produced using massive Landsat time series data (Gong *et al* 2020). This dataset provides annual dynamics of global urban extents at 30-m resolution from 1985 to 2018, using full archives of Landsat images and a consistent mapping approach (Li *et al* 2015, Li and Gong 2016a). The time series impervious area extents in GAIA are temporally consistent (i.e. from non-urban to urban monotonically).

Urban boundary is a fundamental spatial property that can be used to study the impacts of urbanization and other human socio-economic activities on surrounding ecological and environmental phenomena. The definition of urban boundaries varies with different applications and datasets. Population is a commonly used indicator when separating urban and rural areas (Li and Gong 2016a). For example, the urban center (i.e. an approximate boundary) in the Global Human Settlement (GHS) database is defined jointly with population (i.e. more than 1500 inhabitants per km^2) and urban extent data at a 1 km resolution in 2015 (Florczyk *et al* 2019). Thus, the derived urban center in GHS is smaller than the boundary that is commonly used, which includes most artificial impervious areas and associated elements (e.g. parks, lakes, and infrastructures) within the boundary (Jun 2004). NTL observations are widely used to delineate urban boundaries (or extents) from space. The delineation of NTL-based urban boundary is mainly based on the emitted lights from cities at night, which may omit small human settlements with low brightness (Li and Zhou 2017). Additionally, given that there exists a saturation effect in NTL data, the urban extent is mapped using pre-calibrated optimal thresholds, which are determined based on high-resolution land cover maps (Zhou *et al* 2014, 2015). Nonetheless, the GHS and NTL data have relatively coarse resolution (~1 km), resulting in a noticeable saw-toothed effect in derived urban boundaries.

Presently, it is still a great challenge in delineating global urban boundaries from fine resolution urban extent maps in multiple years. Although several attempts have been carried out to delineate urban boundaries from 30-m urban extent maps, most of them were made at the local scale using ancillary datasets or complicated models. Vizzari *et al* (2018) used a clustering approach and land cover data to derive urban boundaries in France. Hu *et al* (2015) integrated land use information entropy model and Kriging interpretation approach to extract the urban boundary in Wuhan, China. Peng *et al* (2016) used a spatial wavelet transform approach to identify the

urban boundary in Beijing from a detailed land use dataset. Taubenböck *et al* (2019) used a sectorial approach to delineate the urban boundary for monocentric cities. The results agree with the general pattern of urban form, whereas they cannot well capture details of urban boundaries in urban fringe areas.

To address these challenges, we built the first global multi-temporal urban boundary dataset using the GAIA data (Gong *et al* 2020). In general, the definition of urban boundaries relies on the high resolution information such as roads, rivers, building blocks, and urban districts (Zhang *et al* 2019), some of which are generally not available at the global scale. Considering the mapping scope being the entire globe with diverse urban environments and multiple years, we adopted the widely used definition of urban boundary, which is mainly based on the spatial distribution of artificial impervious areas from Landsat data (Hu *et al* 2015, Peng *et al* 2016). That is, small urban patches are removed and inner non-urban areas (e.g. green spaces and water bodies) were filled within the boundary of a city (Liang *et al* 2018). To generate such a dataset, we developed an automatic approach to map urban boundaries on the Google Earth Engine (GEE) platform. The remainder of this paper describes the method (section 2), the results with discussion (section 3), and concluding remarks (section 4).

2. Method

We developed a framework to delineate global urban boundaries from the moderate 30 m resolution GAIA data on the GEE platform (figure 1). First, we generated a kernel density map using a kernel density estimation (KDE) approach from 1 km impervious surface area (ISA) data aggregated from GAIA (figure 1(a)). Second, we derived an initial urban boundary using the result from a cellular automata (CA) based modeling approach at a 30 m resolution, combined with the kernel density map from the KDE approach at a 1 km resolution (figure 1(b)). This process can efficiently fill most non-urban areas in/around the urban center. Third, we improved the urban boundary around the urban fringe area, using a morphological approach with dilation and erosion processing (figure 1(c)). Finally, based on the derived urban boundaries, we remove small 'holes' (e.g. small water bodies and green spaces) using a post-processing procedure to retrieve the final boundaries (figure 1(d)). The delineated urban boundary can be used as a basic spatial unit in many applications such as urban planning and assessment of urban development. We implemented this framework on the cloud-based GEE platform (Gorelick *et al* 2017).

2.1. Kernel density map generation

The kernel density map provides an approximate extent to delineate the urban boundary since it smooths the heterogeneity of urban lands in central urban areas (Peng *et al* 2016). First, we produced the ISA data (i.e. the percentage of urban areas within the 1 km grid) by aggregating the 30 m urban extent data. Second, we estimated the kernel density of each ISA pixel using the KDE approach. A symmetric Gaussian point-spread function with a moving circle window of 5 km was used in this step (Zhao *et al* 2019). Compared to the spatial pattern of ISA, the derived kernel density map from the impervious surface areas (ISA-KD) shows a more homogeneous surface in central urban areas (figure S1 (available online at stacks.iop.org/ERL/15/094044/mmedia)).

2.2. Initial urban boundary delineation

The KDE approach fills inner urban areas at a macro-scale, given that the search window is about 25 km^2 (a 5 $\times$ 5 window). We regarded pixels with ISA-KD values above 20% as urban areas (Homer *et al* 2015). This step can fill most inner urban areas (figure S2(a)). However, due to the coarse resolution, the derived urban boundary may miss some small urban patches in/around the city center. Therefore, we filled these inner urban areas using a CA based approach at a 30 m resolution. We used an 11 $\times$ 11 Moore neighborhood in our CA model and treated those pixels with neighborhood densities (i.e. the percentage of urban pixels in the window) above 20% as urban areas (Kocabas and Dragicevic 2006). We iteratively run the urban CA model three times, resulting in a total expansion distance of about 1 km (around 33 pixels) (figure S2(b)). The window size and iteration number were determined under considerations of mapping performance and computation time. Given that the purpose of the CA based approach in this study is to generate the approximate urban boundary, rather than modelling urban expansion, we used the neighborhood component in this CA based approach. Finally, we derived the initial urban boundary by combining results from the urban CA model (30 m) and the KDE approach (1 km) (figure S2(c)). We found that the derived initial urban boundary agrees well with the urban extent from the Google Earth image (figure S2(d)). The combined result from macro- and micro-scales can fill most inner urban areas. Also, the CA based modeling approach can well capture the spatial pattern of urban extents.

2.3. Urban boundary improvement around urban fringe areas

We improved the derived urban boundary around urban fringe areas using a morphological approach. Dilation and erosion are two widely used operations in the set theory (Narayanan 2006). Commonly, these two operations are implemented on a binary image (e.g. urban and non-urban) with a structuring element (e.g. an N $\times$ N window). We chose the structuring element as a 7 $\times$ 7 window as suggested in Liang

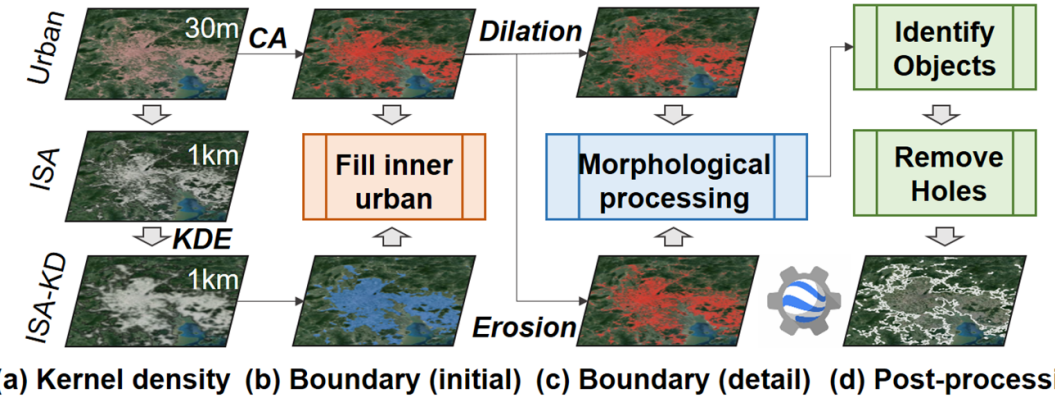

Figure 1. The framework of delineating global urban boundaries from the moderate resolution (30 m) GAIA data, including generating the kernel density map (a), delineating the initial urban boundary (b), improving the urban boundary around the urban fringe area (c), and the post-processing of the derived urban boundaries (d).

et al (2018). The dilation operation can change those non-urban pixels in the structuring element into urban when the structuring element moves around urban pixels. In contrast, the erosion operation can remove those small and isolated urban pixels. We first implemented a dilation operation to combine those small urban patches to the main urban extent (i.e. the initial urban boundary), and then applied an erosion operation to remove isolated urban patches using the same structuring element. By using these two operations sequentially, we improved the urban boundary around urban fringe areas.

2.4. Post-processing

We derived the final urban boundary after a series of post-processing steps (figure 1(d)). First, we identified spatially disconnected urban clusters using an object-based approach. These derived urban clusters are different in terms of their sizes. We screened out small urban clusters (<1 km^2) in this study. Second, we removed holes within urban boundaries of some urban clusters, although the total number of such urban clusters is low. Most of these holes are green spaces and water bodies, which are difficult to be entirely excluded in our previous steps. After these post-processing steps, we delineated global urban boundaries in 1990, 1995, 2000, 2005, 2010, 2015, and the ending year of 2018 in our GAIA data. We did not include 1985 in this study due to the noticeable uncertainty caused by the availability and quality of Landsat data (Gong *et al* 2020). The main urban clusters and its surrounding smaller cities were identified in our approach (figure S3).

3. Results and discussion

3.1. Comparison with other products

Our global urban boundaries (GUB) outperform those from NTL data in terms of the overall pattern of urban boundaries and their spatial details around urban fringe areas (figure 2). The NTL derived urban boundary was mapped from annual observations from the DMSP NTL data (Zhou *et al* 2018). Available results in these two data sources were compared (i.e. 1995, 2000, 2005, and 2010). Overall, urban dynamics reflected by these two data products agree well (figure 2(a)). Compared to the NTL derived result, GUB can provide more details around urban fringe areas (figure 2(b)). Similar results can also be found in other cities (figure S4).

GUB data agree well with the human interpreted results from Landsat images (figure 3), which are commonly used in applications such as urban planning and ecological protection. Here, GUB in China was compared with human interpreted results, that were also based on Landsat images, in three representative years (i.e. 1990, 2000, and 2010) (Wang *et al* 2012). We found GUB agrees well with the interpreted boundaries, not only about the overall pattern (figure 3(a)) but also the spatial details around urban fringe areas (figure 3(b)). However, it should be noted that our approach can: (1) delineate urban boundaries automatically, (2) reduce intensive human labor and subjectivity, and (3) generate consistent results across space and time. The case in Wuhan (China) also suggests that the delineated urban boundary is close to the interpreted result (figure S5).

3.2. Characteristics of GUB

GUB can well capture the spatial extent of urban areas across cities globally (figure 4). Through overlaying the derived boundaries on high resolution Google Earth images, we can find that our mapped results match well with the actual urban extent from satellite images. That is, the boundary can well separate urban and surrounding non-urban areas. In addition, these cities are different in terms of their shapes of derived urban boundaries. For example, cities in North America (e.g. Las Vegas and Edmonton) are more aggregated than cities in Asia (e.g. Bangkok, New Delhi, and Xi'An), of which there are different branches at urban fringe areas. This suggests that the

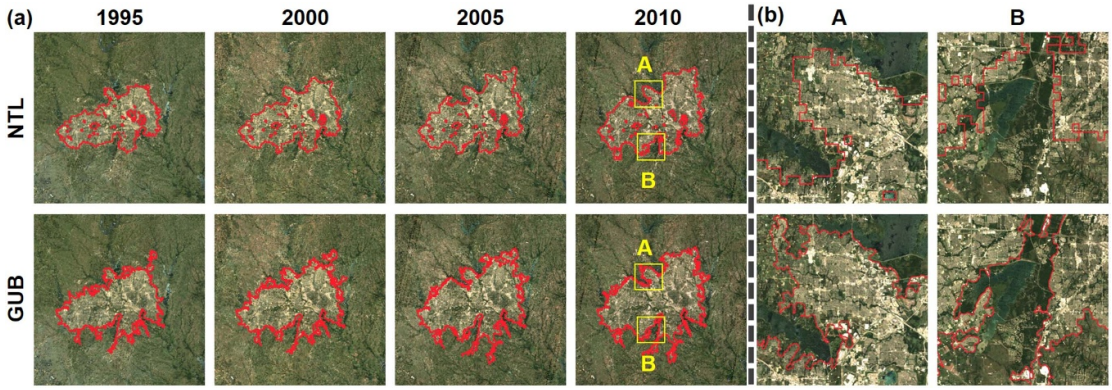

Figure 2. Comparison of urban boundaries derived from GUB and NTL data in Dallas (US) at multiple periods (a) and their enlarged views at the urban fringe areas in 2010 (b). Spatial extents of each subplot in (a) and (b) are 150 km × 150 km and 15 km × 15 km, respectively. Each image is composited by Landsat data using a natural color (R: Band 3, G: Band 2; B: Band 1).

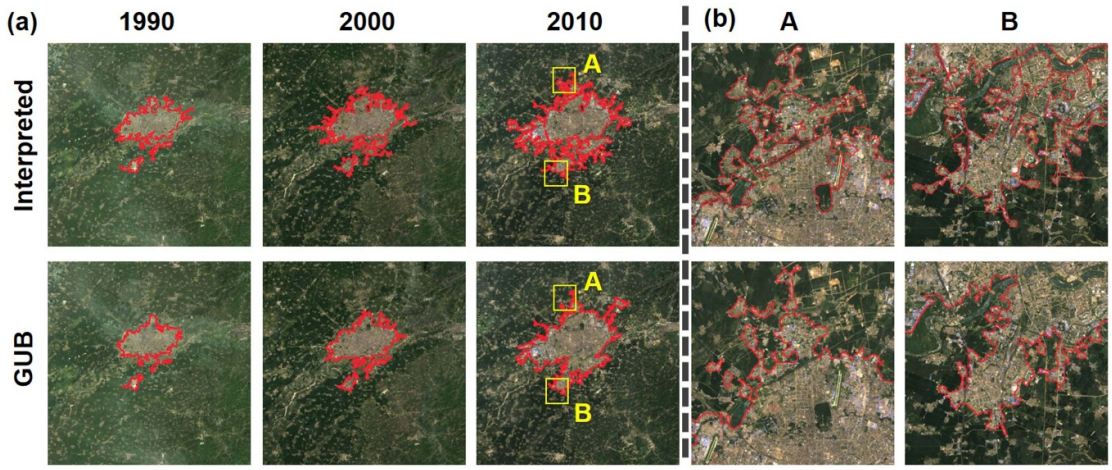

Figure 3. Comparison of urban boundaries from GUB and interpreted results in Shenyang (China) at multiple periods (a) and their enlarged views at the urban fringe areas in 2010 (b). Spatial extents of each subplot in (a) and (b) are 60 km × 60 km and 10 km × 10 km, respectively. Each image is composited by Landsat data using a natural color (R: Band 3, G: Band 2; B: Band 1).

boundaries between urban and rural settlements of cities in Asia are more challenging to identify due to the close interactions between urban and rural areas in these cities (Seto and Kaufmann 2003).

GUB also agrees well with actual urban extents, for both aggregated and scattered cities (figure 5). We selected cities with different distributions and sizes in the US and China for illustration. Given that our approach is consistent at the global scale, aggregated and scattered cities can be well distinguished by analyzing the spatial connectivity between cities and their surrounding urban areas. Aggregated cities with closely connected small settlements were regarded as an entirety and delineated as a large urban cluster (e.g. A1 and A2 in figure 5), while scattered cities isolated from others were delineated as separated urban clusters (e.g. B1, B2, C1, and C2 in figure 5). Although these scatter cities are small, GUB shows a good agreement with the urban extent in high resolution Google Earth images.

In addition, GUB can well capture the dynamics of urban extents (figure 6). The urban extent in Changsha (China) experienced an evident growth over the past three decades. The delineated urban boundaries include most impervious areas and exclude small urban patches around the urban fringe areas (figure 6(a)). This suggests the developed delineation approach works well for the moderate resolution urban extent maps, which have a relatively high degree of heterogeneity in cities. By overlaying the delineated urban boundaries on satellite images, we found the derived results were reliable and matched well with the actual urban extent (figure 6(b)). A similar case can also be found in cities in developed countries (e.g. Des Moines, US, figure S6).

3.3. Urban size comparison with other datasets

The derived urban sizes between GUB data and reference data in the US and China are consistent (figure 7). Urban boundaries in 2010 from the US Census Bureau were digitized with the consideration of population density and land use data within the territory of each census block (https://www.census.gov/).

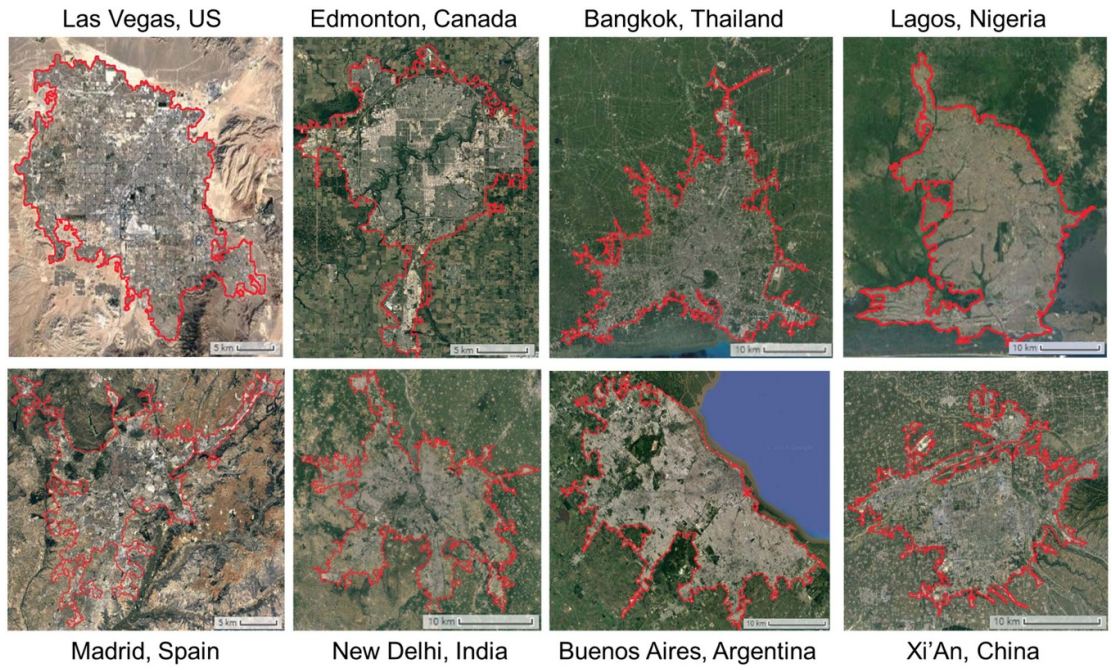

Figure 4. Delineated urban boundaries in eight selected cities in 2018.

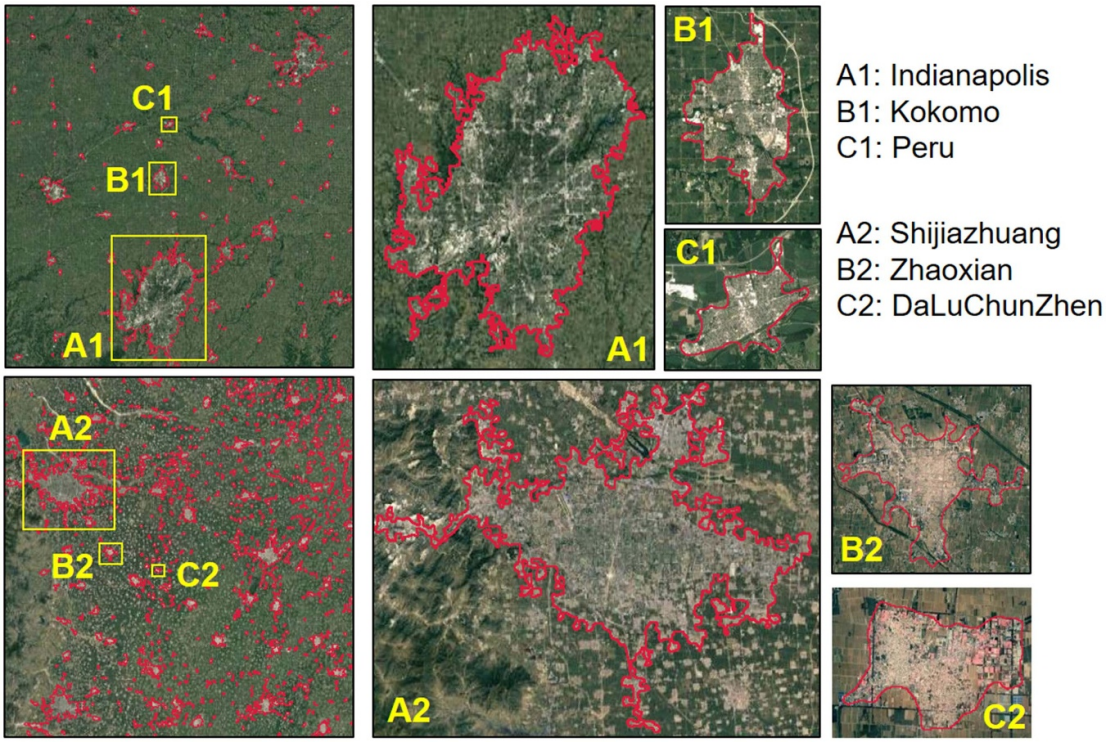

Figure 5. Delineated urban boundaries in example aggregated and scattered cities in the US (upper) and China (bottom).

Urban boundaries from interpreted results in China were generated based on Landsat images in 1990, 2000, and 2010 (Wang *et al* 2012). In general, urban areas from this study and reference datasets are distributed around the 1:1 line (figure 7), with a mean correlation value greater than 0.8. We found urban sizes in our results are slightly smaller than those from the US Census Bureau (figure 7(a)), whereas they are slightly higher than those from the interpreted results in China (figures 7(b)–(d)). This is likely attributed to different definitions and approaches used in delineating urban boundaries in these two reference datasets. While in our results, the mapping approach is consistent over regions and across years. Besides, the

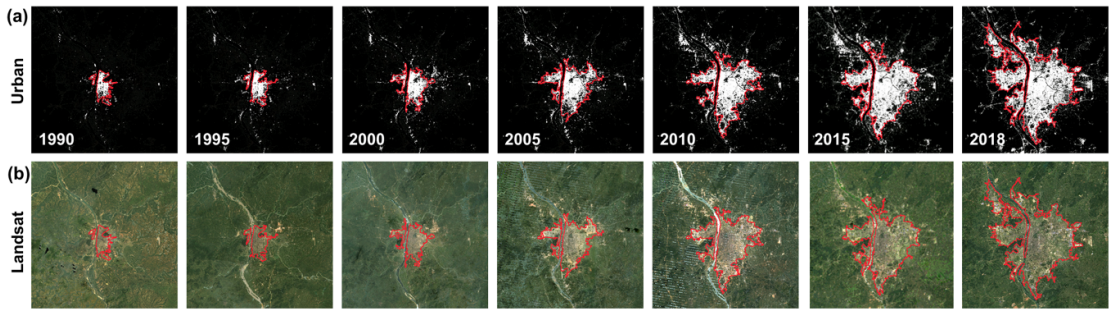

Figure 6. Temporal dynamics of delineated urban boundaries in Changsha (China) on urban extent maps (a) and Landsat images (b). The spatial extent is 60 km × 60 km. Each Landsat image in (b) was composited using a natural color band combination strategy (i.e. R: Band 3, G: Band 2; B: Band 1).

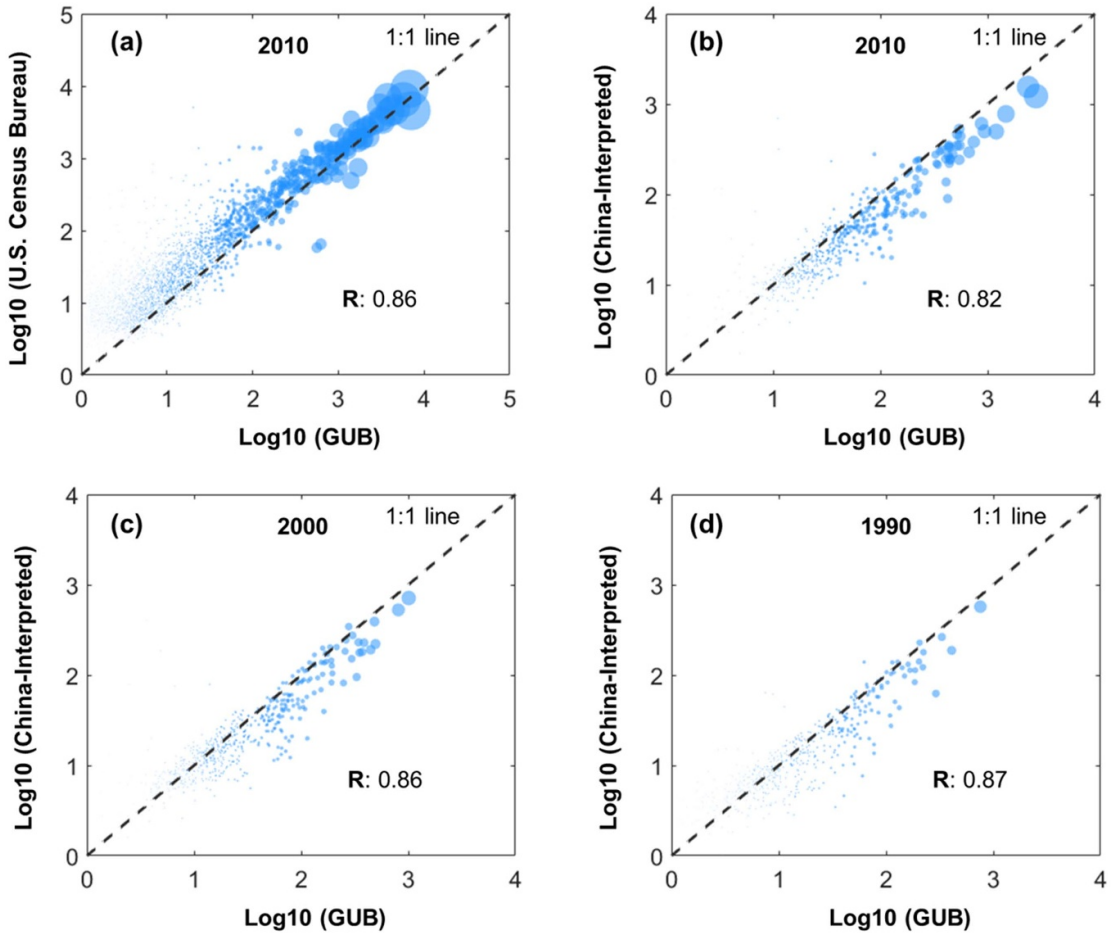

Figure 7. Comparison of urban sizes (km^2) from this study and the reference datasets. The relationship of urban size between GUB and the US. Census Bureau in 2010 (a), and the interpreted results in China in 2010 (b), 2000 (c), and 1990 (d).

correlation between our and interpreted results in China shows a slightly decreasing trend from 1990 to 2010 (i.e. the correlation coefficient decreases from 0.87 to 0.82), indicating the urban landscape in the urban fringe areas becomes more complicated due to the rapid urban expansion.

3.4. Spatiotemporal patterns of urban boundaries across global urban clusters

Currently, most large urban clusters (i.e. >1000 km^2) are distributed in North America, Europe, and coastal areas in China and Japan, but they show different growth trends over the past three decades (figure 8(a)). In general, urban clusters in developed regions (e.g. the US and Europe) show larger sizes than those in developed regions such as India and China. This is likely attributed to their varying urbanization levels, resulting in different spatial patterns of human settlements in urban and rural areas. For example, the urbanization level (i.e. percentage of urban population) is above 80% in the US, higher than those in China (55%) and India (30%) (United Nations 2019).

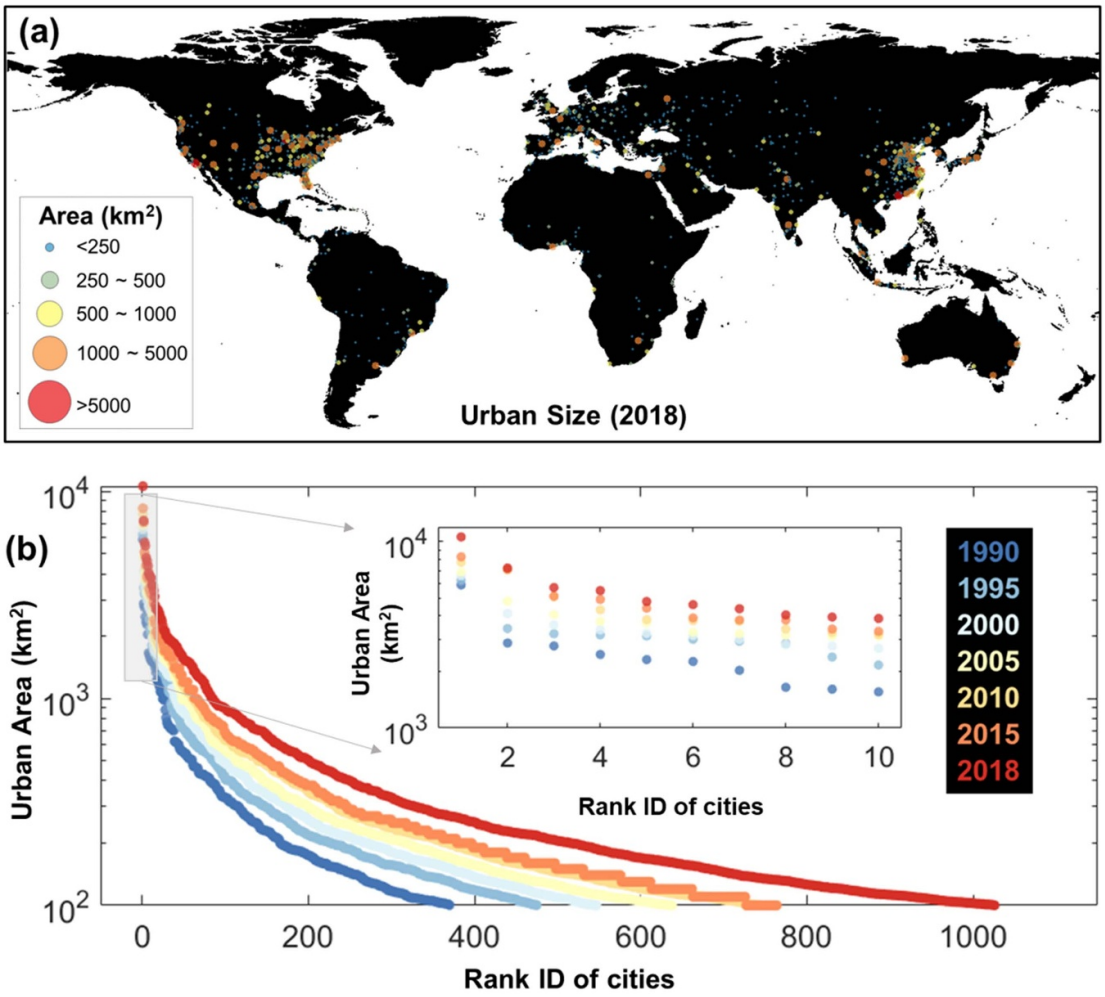

Figure 8. Distribution of global urban clusters (greater than 100 km^2) in 2018 (a) and their temporal dynamics of size-rank relationship over the past three decades (b). The top 10 urban clusters in 1990 are Los Angeles (US), Dallas (US), Tokyo (Japan), Chicago (US), Johannesburg (South Africa), Miami (US), Detroit (US), New York (US), Bay Area (US), and Seattle (US). The top 10 urban clusters in 2018 are Pearl River Delta (China), Los Angeles (US), Yangtze River Delta (China), Chicago (US), Dallas (US), Houston (US), Beijing (China), Johannesburg (South Africa), New York (US), and Shanghai (China).

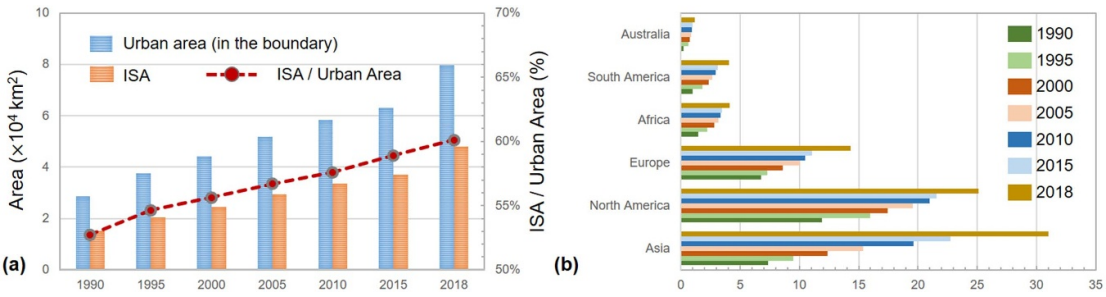

Figure 9. Change of urban areas within delineated boundaries. Comparison of urban areas and impervious surface areas (ISA) at the global (a) and continental scales (b).

As a result, there are many small human settlements in China and India, although their total numbers of large urban clusters are relatively low.

In addition, we observed a notable expansion of urban areas over the past three decades (figure 8(b)). The total number of large urban clusters in 1990 is about 400, and this number is more than doubled (1000) in 2018 (figure 8(a)). Curves of the urban area and city rank also clearly illustrate a considerable growth in terms of the size and number of large urban clusters. On the one hand, medium-size urban clusters with the same rank order from 1990 to 2018 show a notable increase in urban areas. On the other hand, for those large urban clusters, differences of their areas are enlarged with the increase in the rank order and years, suggesting that those large cities also experienced a noticeable increase. Besides, some large urban clusters in developed countries (e.g. Seattle,

Detroit, and Tokyo) were replaced by several rapidly developing urban clusters in Asia (e.g. Yangtze river delta and Pearl river delta). It should be noted that the formation of these mega-cities is likely related to the expansion of cities and their merging with nearby small cities (Hu *et al* 2019). Although such a merging process will reduce the number of urban clusters, the total number of urban clusters in more than 90% of countries is increasing over the past decades due to the growth of massive small urban clusters. For example, China and the US account for around 40% of urban clusters in the world, and their increases in numbers of urban clusters (1990–2018) are from 4725 to 13 916 and from 7426 to 11 337, respectively. However, for some countries without notable growth of small urban clusters (e.g. Argentina), their numbers of urban clusters decrease due to the merging of surrounding small cities. Details of cluster numbers in different countries can be found in supplementary table 1.

Globally, the total urban area in delineated boundaries for urban clusters greater than 1 km^2 is 809 665 km^2 in 2018, of which around 60% are impervious areas (figure 9). Although the global urban area within GUB is close to the ISA (~800 000 km^2) in GAIA in 2018, the areas in GUB include non-impervious areas inside the delineated boundaries but exclude small and diffused urban areas outside the boundaries (figure 9(a)). At the global scale, the proportion of ISA in delineated boundaries ranges from 53% to 60%, showing a consistently increasing trend from 1990 to 2018. This is likely attributed to the notable urban expansion over past decades since this proportion grows with the increase of urban size (e.g. non-urban regions within the boundary become urban due to urban sprawl) (figure S7). Around 90% of urban areas within the delineated boundaries are in Asia, North America, and Europe. Asia alone occupies around 40% of the total area in the world (figure 9(b)). Proportions of ISA within delineated boundaries are different across continents (figure S8(a)), despite that their trends are increasing over past decades. Overall, this proportion is relatively high in Africa and low in Europe among these continents. The relatively small urban cluster size in Europe contributes to the low proportion of ISA in delineated boundaries (figure S8(b)), similar to results in figure S7. Also, urban environments of cities in these two continents are notably different, where impervious areas within the boundaries are more compact in Africa than Europe (figure 9).

3.5. Implication of per capita urban area

Among the top 10 most urbanized nations in 2018, the US has the highest per capita urban areas (i.e. more than 800 m^2). Within the delineated boundaries, we measured the impervious areas from the GAIA data (Gong *et al* 2020) and the population from the LandScan (Dobson *et al* 2000) (supplementary table 2). The top 10 most urbanized nations in 2018 are the US, China, India, Russia, Brazil, Japan, France, Germany, Canada, and Italy, while the top 10 countries in terms of their per capita urban areas are the US, Canada, France, Italy, Germany, Russia, China, Brazil, Japan, and India. As the top two countries of urban areas, urban area in the US is almost 1.2 times of that in China, while the per capita urban area in the US is around tripled (2.7 times) that in China. Besides, Japan has the largest proportion (70%) of ISA within delineated boundaries in these top 10 countries, due to the relatively high population density in cities. Details of country-based urban areas, impervious surface areas, population, and per capita urban areas can be found in supplementary table 2.

4. Conclusions

In this study, we developed a framework for delineating urban boundaries from the GAIA data and thereby generated the multi-temporal GUB data on the GEE platform in 1990, 1995, 2000, 2005, 2010, 2015, and 2018. First, we delineated the initial urban boundary for urban clusters through filling inner non-urban areas, using a combined approach of KDE and CA based urban growth modeling. Second, we improved the initially derived urban boundaries in fringe areas using a morphological approach.

The GUB dataset shows a good agreement with results from other products. The comparison with NTL derived results suggests the delineated urban boundaries in GUB can well capture the complicated shapes (or geometries) of urban extent around urban fringe areas. Such improvements in our delineation approach can result in similar urban boundaries as those obtained from human interpretation. Also, the change in GUB is consistent with the dynamics of built-up areas as reflected by the Landsat time series data. Globally, the total urban area in our GUB dataset is more than 800 000 km^2 in 2018, about 40% greater than the impervious surface area in GAIA.

This dataset shows potentials in various urbanization related studies such as sustainable planning and ecological protection. Different from other products (e.g. the NTL derived urban extent), our results are advanced in its capacity of delineating detailed urban boundaries around the urban fringe areas. These peri-urban areas are also the most sensitive regions due to the rapid urban expansion and vulnerability of biodiversity and agricultural ecosystems (Seto *et al* 2014). Therefore, our datasets can support a variety of urban studies, such as the impact of urbanization on the urban ecosystem, urban landscape, and urban climate (Mcdonald *et al* 2019). The GUB dataset, including results before and after post-processing (e.g. figure S9), can be freely downloaded from http://data.ess.tsinghua.edu.cn. We will update the GAIA and GUB datasets in the future.

Acknowledgments

This work is partly supported by the National Key Research and Development Plan of the People's Republic of China (No. 2019YFA0607201 and No. 2017YFA0604404), donations from Delos Living LLC and the Cyrus Tang Foundation to Tsinghua University, and the College of Liberal Arts and Science's (LAS) Dean's Emerging Faculty Leaders award at the Iowa State University.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Alberti M *et al* 2017 Global urban signatures of phenotypic change in animal and plant populations *Proc. Nat. Acad. Sci.* **114** 8951–6

Clinton N and Gong P 2013 MODIS detected surface urban heat islands and sinks: global locations and controls *Remote Sens. Environ.* **134** 294–304

DeFries R, Rudel T, Uriarte M and Hansen M 2010 Deforestation driven by urban population growth and agricultural trade in the twenty-first century *Nat. Geosci.* **3** 178–81

Dobson J E *et al* 2000 LandScan: a global population database for estimating populations at risk *Photogramm. Eng. Remote Sens.* **66** 849–57

Elvidge C D, Tuttle B, Sutton P, Baugh K, Howard A, Milesi C, Bhaduri B and Nemani R 2007 Global distribution and density of constructed impervious surfaces *Sensors* **7** 1962–79

Esch T, Marconcini M, Felbier A, Roth A, Heldens W, Huber M, Schwinger M, Taubenbock H, Muller A and Dech S 2013 Urban footprint processor—fully automated processing chain generating settlement masks from global data of the TanDEM-X mission *IEEE Geosci. Remote Sens. Lett.* **10** 1617–21

Florczyk A *et al* 2019 Description of the GHS Urban Centre Database 2015 Publications Office of the European Union http://doi.org/10.2760/037310

Foley J *et al* 2005 Global consequences of land use *Science* **309** 570–4

Foley J *et al* 2011 Solutions for a cultivated planet *Nature* **478** 337–42

Gong P *et al* 2020 Annual maps of global artificial impervious areas (GAIA) between 1985 and 2018 *Remote Sens. Environ.* **236** 111510

Gong P, Li X and Zhang W 2019a 40-year (1978-2017) human settlement changes in China reflected by impervious surfaces from satellite remote sensing *Sci. Bull.* **64** 756–63

Gong P, Liang S, Carlton E J, Jiang Q, Wu J, Wang L and Remais J V 2012 Urbanisation and health in China *Lancet* **379** 843–52

Gong P *et al* 2019b Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017 *Sci. Bull.* **64** 370–3

Gong P *et al* 2013 Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data *Int. J. Remote Sens.* **34** 2607–54

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D and Moore R 2017 Google Earth Engine: planetary-scale geospatial analysis for everyone *Remote Sens. Environ.* **202** 18–27

Güneralp B *et al* 2017 Global scenarios of urban density and its impacts on building energy use through 2050 *Proc. Nat. Acad. Sci.* 114 8945–50

Homer C G *et al* 2015 Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information *Photogramm. Eng. Remote Sens.* **81** 345–54

Hu S, Tong L, Frazier A E and Liu Y 2015 Urban boundary extraction and sprawl analysis using Landsat images: a case study in Wuhan, China *Habitat Int.* **47** 183–95

Hu T *et al* 2019 Toward a better understanding of urban sprawl: linking spatial metrics and landscape networks dynamics *Int. Conf. on Computers in Urban Planning and Urban Management* (Berlin: Springer) pp 163–78

Jun M-J 2004 The effects of Portland's urban growth boundary on urban development patterns and commuting *Urban Stud.* **41** 1333–48

Kocabas V and Dragicevic S 2006 Assessing cellular automata model behaviour using a sensitivity analysis approach *Comput. Environ. Urban Syst.* **30** 921–53

Li X and Gong P 2016a An "exclusion-inclusion" framework for extracting human settlements in rapidly developing regions of China from Landsat images *Remote Sens. Environ.* **186** 286–96

Li X and Gong P 2016b Urban growth models: progress and perspective *Sci. Bull.* **61** 1637–50

Li X, Gong P and Liang L 2015 A 30-year (1984–2013) record of annual urban dynamics of Beijing City derived from Landsat data *Remote Sens. Environ.* **166** 78–90

Li X, Yu L, Xu Y, Yang J and Gong P 2016 Ten years after Hurricane Katrina: monitoring recovery in New Orleans and the surrounding areas using remote sensing *Sci. Bull.* **61** 1460–70

Li X and Zhou Y 2017 Urban mapping using DMSP/OLS stable night-time light: a review *Int. J. Remote Sens.* **38** 1–17

Li X, Zhou Y, Asrar G R, Mao J, Li X and Li W 2017 Response of vegetation phenology to urbanization in the conterminous United States *Glob. Change Biol.* **23** 2818–30

Li X, Zhou Y, Eom J, Yu S and Asrar G R 2019 Projecting global urban area growth through 2100 based on historical time-series data and future Shared Socioeconomic Pathways *Earth's Future* **7** 351–62

Li X, Zhou Y, Gong P, Seto K C and Clinton N 2020a Developing a method to estimate building height from Sentinel-1 data *Remote Sens. Environ.* **240** 111705

Li X, Zhou Y, Zhu Z and Cao W 2020b A national dataset of 30 m annual urban extent dynamics (1985–2015) in the conterminous United States *Earth Syst. Sci. Data* **12** 357–71

Li X, Zhou Y, Zhu Z, Liang L, Yu B and Cao W 2018 Mapping annual urban dynamics (1985-2015) using time series of Landsat data *Remote Sens. Environ.* **216** 674–83

Liang X, Liu X, Li X, Chen Y, Tian H and Yao Y 2018 Delineating multi-scenario urban growth boundaries with a CA-based FLUS model and morphological method *Landsc Urban Plan.* **177** 47–63

Liu X, Hu G, Chen Y, Li X, Xu X, Li S, Pei F and Wang S 2018 High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform *Remote Sens. Environ.* **209** 227–39

Liu Z, He C, Zhou Y and Wu J 2014 How much of the world's land has been urbanized, really? A hierarchical framework for avoiding confusion *Landsc. Ecol.* **29** 763–71

Lu Y, Nakicenovic N, Visbeck M and Stevance A-S 2015 Five priorities for the UN sustainable development goals *Nature* **520** 432–3

Mbow H-O P *et al* 2017 Special Report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems (SR2) (www.ipcc.ch/site/assets/uploads/2018/07/sr2_background_report_final.pdf)

Mcdonald R I *et al* 2019 Research gaps in knowledge of the impact of urban growth on biodiversity *Nat. Sustain.* **3** 16–24

Narayanan A 2006 Fast binary dilation/erosion algorithm using kernel subdivision *Asian Conf. on Computer Vision* (Berlin: Springer) pp 335–42

Peng J, Zhao S, Liu Y and Tian L 2016 Identifying the urban-rural fringe using wavelet transform and kernel density estimation: a case study in Beijing City, China *Environ. Model. Softw.* **83** 286–302

Pesaresi M *et al* 2015 GHS built-up grid, derived from Landsat, multitemporal (1975, 1990, 2000, 2014) (https://ec.europa.eu/jrc/en/publication/ghs-built-grid-derived-landsat-multitemporal-1975-1990-2000-2014-ir2017-v10)

Ratcliffe M *et al* 2016 Defining rural at the US Census Bureau *American Community Survey and Geography Brief* pp 1–8 (https://www.researchgate.net/profile/Michael_Ratcliffe2/publication/311533270_Defining_Rural_at_the_US_Census_Bureau/links/584aad3708aeb19dcb758910/Defining-Rural-at-the-US-Census-Bureau.pdf)

Schneider A, Friedl M A and Potere D 2010 Mapping global urban areas using MODIS 500-m data: new methods and datasets based on 'urban ecoregions' *Remote Sens. Environ.* **114** 1733–46

Seto K C *et al* 2014 Human settlements, infrastructure and spatial planning *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* ed O Edenhofer *et al* (Cambridge: Cambridge University Press)

Seto K C and Kaufmann R K 2003 Modeling the drivers of urban land use change in the Pearl River Delta, China: integrating remote sensing with socioeconomic data *Land Econ.* **79** 106–21

Taubenböck H, Weigand M, Esch T, Staab J, Wurm M, Mast J and Dech S 2019 A new ranking of the world's largest cities—do administrative units obscure morphological realities? *Remote Sens. Environ.* **232** 111353

United Nations 2019 *World Urbanization Prospects: The 2018 Revision* (UN) (https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf)

Vizzari M, Hilal M, Sigura M, Antognelli S and Joly D 2018 Urban-rural-natural gradient analysis with CORINE data: an application to the metropolitan France *Landsc. Urban Plan.* **171** 18–29

Wang L *et al* 2012 China's urban expansion from 1990 to 2010 determined with satellite remote sensing *Chin. Sci. Bull.* **57** 2802–12

Wang S, Ju W, Peñuelas J, Cescatti A, Zhou Y, Fu Y, Huete A, Liu M and Zhang Y 2019 Urban− rural gradients reveal joint control of elevated CO_2 and temperature on extended photosynthetic seasons *Nat. Ecol. Evol.* **3** 1076–85

Xi F *et al* 2016 Substantial global carbon uptake by cement carbonation *Nat. Geosci.* **9** 880–3

Zhang H, Ning X, Shao Z and Wang H 2019 Spatiotemporal pattern analysis of China's cities based on high-resolution imagery from 2000 to 2015 *ISPRS Int. J. Geo-Inf.* **8** 241

Zhang Q, He K and Huo H 2012 Policy: cleaning China's air *Nature* **484** 161–2

Zhang W, Villarini G, Vecchi G A and Smith J A 2018 Urbanization exacerbated the rainfall and flooding caused by hurricane Harvey in Houston *Nature* **563** 384

Zhao M, Zhou Y, Li X, Zhou C, Cheng W, Li M and Huang K 2019 Building a series of consistent night-time light data (1992-2018) in southeast Asia by integrating DMSP-OLS and NPP-VIIRS *IEEE Trans. Geosci. Remote Sens.* **58** 1843–56

Zhou L *et al* 2004 Evidence for a significant urbanization effect on climate in China *Proc. Nat. Acad. Sci. USA* **101** 9540–4

Zhou Y, Li X, Asrar G R, Smith S J and Imhoff M 2018 A global record of annual urban dynamics (1992-2013) from nighttime lights *Remote Sens. Environ.* **219** 206–20

Zhou Y, Smith S J, Elvidge C D, Zhao K, Thomson A and Imhoff M 2014 A cluster-based method to map urban area from DMSP/OLS nightlights *Remote Sens. Environ.* **147** 173–85

Zhou Y, Smith S J, Zhao K, Imhoff M, Thomson A, Bond-Lamberty B, Asrar G R, Zhang X, He C and Elvidge C D 2015 A global map of urban extent from nightlights *Environ. Res. Lett.* **10** 1–11