Remote Sensing of Environment 147 (2014) 173–185

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

A cluster-based method to map urban area from DMSP/OLS nightlights

Yuyu Zhou [a,*], Steven J. Smith [a], Christopher D. Elvidge [b], Kaiguang Zhao [c], Allison Thomson [a], Marc Imhoff [a]

[a] Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD 20740, United States
[b] Earth Observation Group, NOAA National Geophysical Data Center, Boulder, CO 80303, United States
[c] School of Environment and Natural Resources, The Ohio State University, Wooster, OH 44691, United States

ARTICLE INFO

Article history:
Received 5 November 2013
Received in revised form 25 February 2014
Accepted 1 March 2014
Available online xxxx

Keywords:
DMSP/OLS
Nightlights
Urban area
Threshold
Cities
Segmentation
Land cover and land use change

ABSTRACT

Accurate information on urban areas at regional and global scales is important for both the science and policy-making communities. The Defense Meteorological Satellite Program/Operational Linescan System (DMSP/OLS) nighttime stable light data (NTL) provide a potential way to map the extent and dynamics of urban areas in an economic and timely manner. In this study, we developed a cluster-based method to estimate optimal thresholds and map urban extent from the DMSP/OLS NTL data in five major steps, including data preprocessing, urban cluster segmentation, logistic model development, threshold estimation, and urban extent delineation. In our method the optimal thresholds vary by clusters and are estimated based on cluster size and overall nightlight magnitude. The United States and China, two large countries with different urbanization patterns, were selected to test the proposed method. Our results indicate that the urbanized area occupies about 2% of total land area in the US, ranging from lower than 0.5% to higher than 10% at the state level, and less than 1% in China, ranging from lower than 0.1% to about 5% at the province level with some municipalities as high as 10%. The derived thresholds and urban extent were evaluated using a validation sub-sample of high-resolution land cover data at the cluster and regional levels. It was found that our method can map urban areas in both countries efficiently and accurately. The sensitivity analysis indicates that the derived optimal thresholds are not highly sensitive to the parameter choices in the logistic model. Our method reduces the over- and under-estimation issues often associated with previous fixed-threshold techniques when mapping urban extent over a large area. More importantly, our method shows potential to map global urban extent and temporal dynamics using the DMSP/OLS NTL data in a timely, cost-effective way.

1. Introduction

The urban system is complex with various interacting components. Urbanized area, a major feature of the urban system, represents population centers and economic hubs largely characterized by surfaces occupied by buildings, streets, and other infrastructure (Zhang & Seto, 2011). Although urban areas occupy a relatively small fraction of total Earth's surface, urbanization is one of the most important components of human-induced land cover and land use change (LCLUC) and has profound impacts on energy (e.g. urban heat island), water (e.g. flooding), pollution, ecosystems, and carbon cycle from local to regional and even global scales (Brabec, 2002; Foley et al., 2005; McKinney, 2008; Shepherd, 2005; Zhou, Wang, Gold, & August, 2010; Zhou, Wang, Gold, August, & Boving, 2013; Zhou, Weng, Gurney, Shuai, & Hu, 2012). For example, a previous study indicated that 37–86% of direct fuel consumption in buildings and industry and 37–77% of on-road gasoline and diesel consumption in the US occurred in urban areas (Parshall et al., 2010) and the placement of urban infrastructure, while small in area, has a disproportionate impact on potential net primary productivity because of the high native fertility of transformed soils (Imhoff et al., 2004; Nizeyimana et al., 2001).

Remote sensing has been recognized as a major source of consistent and continuous data, and has been used to study urbanization and its change across a variety of temporal and spatial scales (Schneider et al., 2010; Zhang & Seto, 2011; Zhou & Wang, 2007; Zhou & Wang, 2008). Much progress has been made in urbanization research using remote sensing in terms of methodology development and analysis. Urbanization and its related dynamics have been studied not only for individual cities or greater metropolitan areas, but also across selected cities for comparative purposes (Schneider & Woodcock, 2008). Although researchers have started to pay attention to urbanization over large areas, even at global scales (Zhang & Seto, 2011), there are still limited investigations of large scale urban dynamics primarily due to the lack of efficient and timely methods for mapping urban extent over large areas.

Moderate spatial resolution remote sensing data have demonstrated their capability in large scale and global urbanization mapping (Elvidge, Sutton, et al., 2009; Elvidge, Tuttle, et al., 2007; Loveland et al., 2000;

* Corresponding author. Tel.: +1 301 314 6771.
E-mail address: Yuyu.zhou@pnnl.gov (Y. Zhou).

http://dx.doi.org/10.1016/j.rse.2014.03.004
0034-4257/

Schneider et al., 2010). For example, Schneider et al. (2010) developed a 500 m resolution global urban map using MODIS data from 2000 to 2002. Elvidge, Safran et al. (2007) built a global impervious surface areas (ISA) map using nighttime lights, population counts, and high-resolution ISA data. European Space Agency (ESA) generated a global land cover map using the 300 m Medium Resolution Imaging Spectrometer (MERIS) time series dataset (ESA, 2013). Moreover, with the help of other data and techniques, a number of global urban or population distribution maps have been developed, i.e. the LandScan product (Dobson, Bright, Coleman, Durfee, & Worley, 2000) and the Global Rural–Urban Mapping Project (GRUMP) urban extent (CIESIN, 2011). However, most of these global products have limited temporal coverage, with limited usefulness for dynamic analysis at large scales. Although urban density (fractional urbanization) maps, e.g. global map of ISA by Elvidge, Safran, et al. (2007) and Elvidge, Tuttle, et al. (2007), can provide more information for the study of urbanization, these products require further information such as population or higher resolution supplementary data, which may be difficult to obtain for long time periods over large scales. Moreover, some of these methods require labor-intensive processing of a sufficient number of cloud-free images, and issues of spectral and spatial consistency from different scenes may exist.

The Defense Meteorological Satellite Program/Operational Linescan System (DMSP/OLS) nighttime stable light data (NTL) data are, therefore, a valuable resource for regional and global urban mapping and application to the study of human activities such as population density, economic activity, energy use, and CO_2 emissions (Amaral, Câmara, Monteiro, Quintanilha, & Elvidge, 2005; Cao, Chen, Imura, & Higashi, 2009; Doll, Muller, & Elvidge, 2000; Elvidge, Baugh, Kihn, Kroehl, Davis, et al., 1997; Elvidge, Safran, et al., 2007; Elvidge, Tuttle, et al., 2007; Imhoff, Lawrence, Stutzer, & Elvidge, 1997; Oda & Maksyutov, 2011; Sutton, 2003; Zhang & Seto, 2011). However, there are several shortcomings in this data, including limited dynamic range, signal saturation in urban centers, contamination from other sources such as gas flares, lack of a well-characterized point spread function (PSF), and lack of a well-characterized field of view (Elvidge, Sutton, et al., 2009). In particular, OLS-derived light features are substantially larger than the lighting sources on the ground, and local economic conditions may have different impacts on the detection and brightness of satellite observed lighting (Elvidge, Sutton, et al., 2009). It was found that the DMSP/OLS NTL data tend to exaggerate the size of urban areas compared to the Landsat analysis, due to several contributing factors, including the reflectance of light from surrounding water and non-urban land areas, georeferencing errors, and warm atmospheric phenomena (Henderson, Yeh, Gong, Elvidge, & Baugh, 2003). For example, the considerable excursion of reflected light onto water bodies causes pixel blooming along the shorelines of large metropolitan areas and the resulting overestimation produces enlarged small towns and expanded boundaries of large cities (Imhoff et al., 1997).

A number of methods have been developed to map urban areas using the DMSP/OLS NTL data (Cao et al., 2009; Elvidge, Tuttle, et al., 2007; Frolking, Milliman, Seto, & Friedl, 2013; He et al., 2006; Liu, He, Zhang, Huang, & Yang, 2012; Lu, Tian, Zhou, & Ge, 2008; Owen, 1998; Small, Pozzi, & Elvidge, 2005; Sutton, Cova, & Elvidge, 2006). Simple threshold techniques showed potential in generating reasonable urban mapping products at the regional and national scales by using the DMSP/OLS NTL data (Amaral et al., 2005; Henderson et al., 2003; Imhoff et al., 1997; Kasimu, Tateishi, & Hoan, 2009). However, the choices of optimal thresholds may vary across regions and countries due to the regional variation in physical environment and socioeconomic development status (Cao et al., 2009; Liu et al., 2012; Small et al., 2005). The determination of appropriate thresholds in delineating urban areas using the DMSP/OLS NTL data is one of the major challenges in urban mapping over large areas (Henderson et al., 2003). Applying a single threshold to the DMSP/OLS data may be problematic, especially across multiple cities or political boundaries (Imhoff et al., 1997). For example, Henderson et al. (2003) found a range in optimal thresholds for urban mapping across different cities with stable light land area lit thresholds of 92% for San Francisco, 97% for Beijing, and 88% for Lhasa, all of which were higher than the thresholds of 82% and 89% for the continental US reported by Imhoff et al. (1997).

Due to the issues in existing global and regional based threshold techniques and their inflexibility, it is necessary, and also a research challenge, to derive optimal thresholds specific to different cities or urban clusters using the DMSP/OLS NTL data in ways that are neither costly nor complex and are globally applicable. In this study, we developed a cluster-based method to estimate the optimal thresholds and delineate the urban extent, and selected the contiguous United States and China, two countries with different urbanization patterns, and also with high quality land-cover data, as experimental areas. This paper focuses on the development of the new threshold method through calibration and validation using a sub-set of regional high-resolution reference data. The remainder of this paper describes the study area and data (Section 2), details of the five major steps of our method (Section 3), a discussion of the results and findings (Section 4), and concluding remarks (Section 5).

2. Study area and data

In this study, the contiguous US and China were chosen as the experimental areas. These two study areas have different urbanization patterns. In particular, urbanization levels in China vary greatly across space, attributable to the heterogeneous socioeconomic development whereas urbanization is somewhat more uniform in the US. The different urbanization densities and patterns in the US and China provide ideal experimental regions for evaluating the global applicability of the proposed urban mapping method.

The major data used in this study are DMSP/OLS NTL, high spatial resolution regional land cover, a water mask, and a gas flare mask. The DMSP/OLS, designed to collect global cloud imagery (Elvidge, Erwin, et al., 2009), can provide a systematically collected, unbiased global nightlight dataset, and has a number of unique features that meet the needs of wide-scale, frequently repeated surveys of urban growth (Henderson et al., 2003). More importantly, the DMSP/OLS NTL data have an annual temporal coverage at the global level from 1992 to the present. The DMSP/OLS NTL measures lights on the Earth's surface from cities and settlements with persistent lighting, and others such as gas flares, fires, and illuminated marine vessels (Zhang, Schaaf, & Seto, 2013). The data at each pixel are recorded as a digital number (DN) from 0 to 63 with a 1 km spatial resolution, spanning −180° to +180° in longitude and −65° to +75° in latitude. The annual cloud-free composites were built using the highest-quality data based on a number of constraints (Elvidge, Zisken, et al., 2009). In this study, we chose NTL data in the years 2006 and 2005 for the US and China, respectively, to be temporally consistent with the high spatial resolution land cover datasets used for training and evaluation.

High spatial resolution land cover datasets were acquired from existing sources for developing and testing the proposed method. Specifically, the high-resolution data for the US and China were obtained from the US Geological Survey National Land Cover Dataset (NLCD) and the Resources and Environment Data Center of the Chinese Academy of Science, respectively, both with an original spatial resolution of 30 m (Homer, Huang, Yang, Wylie, & Coan, 2004; Liu et al., 2010). The land cover types mainly include open water, urban, evergreen forest, deciduous forest, shrub, grassland, cropland and wetland. The land-cover data for China were built through visual interpretation of Landsat TM images and processed to a 1 km percentage map of each land cover type (Liu et al., 2010). The US land-cover data layer was also upscaled from a 30 m to a 1 km spatial resolution. Urban areas from all 30 m pixels within a 1 km pixel were summed and converted to percentage, resulting in an urban percentage map. To be consistent with the binary urban map we will derive from the nightlights data, we need to

translate percentage urban cover from the land-use data to a binary classification. Here, we define urban land as those 1 km pixels with urban percentages larger than 20%, which is consistent with the land use category of developed areas (impervious surfaces ≥20%) as defined in the NLCD dataset (Fry et al., 2011). Thus, one 1 km binary urban map was constructed for each country considered. In addition, a water mask was derived from MODIS 250 m land-water mask (MOD44W), and gas flare data were obtained from the NOAA National Geographic Data Center (Elvidge, Ziskin, et al., 2009).

3. Methods

Threshold techniques have shown potential in generating reasonable urban mapping products at the regional and national scales by using the DMSP/OLS NTL data. In these methods, the pixels with NTL magnitude values larger than some optimal threshold value are classified as urban and all others as nonurban. However, determining optimal thresholds for all cities in a study region is difficult and still remains a challenge. We develop here a cluster-based method to estimate optimal thresholds and map urban extent using DMSP/OLS NTL and supplementary data. This method includes five major steps: data preprocessing, urban cluster segmentation, logistic model development, threshold estimation, and urban extent delineation (Fig. 1). First, we filtered NTL data by excluding water and gas flare pixels. Second, we identified potential urban clusters, which are groups of similar and continuous pixels, from filtered NTL data using a segmentation method. Third, we analyzed the relationship of optimal threshold derived from high spatial resolution land cover data with cluster size and NTL magnitude in each cluster, and built a logistic model for optimal threshold estimation. In our method, the threshold is defined as the DMSP/OLS DN value above which the pixel is classified as urban area. The threshold value is allowed to vary between urban clusters. Fourth, we estimated the optimal threshold value for each cluster using the logistic model. Finally, we mapped the urban extent according to the estimated threshold in each cluster. Each step will be discussed in detail in the following subsections.

3.1. Data preprocessing

In the data preprocessing step, we filtered the original NTL data by removal of water and gas flare pixels. All of the data were processed to the same projection and spatial resolution as those of the DMSP/OLS NTL data. Gas flares have features similar to urban lighting in the DMSP/OLS NTL data, and they occur in remote locations, outside urban centers (Elvidge, Ziskin, et al., 2009). They introduce errors in urban mapping and may cause an overestimation of the urban areas. The gas flare data were produced from DMSP/OLS NTL data by Elvidge, Ziskin, et al. (2009). For this product, a 30 arc second global population density grid from the US DOE was used to evaluate lights identified as potential gas flares. NASA MODIS satellite hot spot data were also used to assist in clarifying the identity of gas flares on land. A mask of gas flare pixels is used to exclude these pixels to eliminate this element of overestimation. A water mask from the MODIS MOD44W product was used to exclude pixels with water percentage larger than 50% to reduce the influence of pixel blooming along the shorelines of large metropolitan areas.

3.2. Urban cluster segmentation

The next step was to delineate potential urban clusters, which are areas composed of continuous similarly lighted areas in NTL images. This application requires an automatic and efficient method that is effective for single-band NTL DN values. The Marker-controlled Watershed Segmentation algorithm was chosen to segment the filtered NTL image after gas flares and water pixels were excluded (Parvati et al., 2008). A number of segmentation algorithms have been developed,

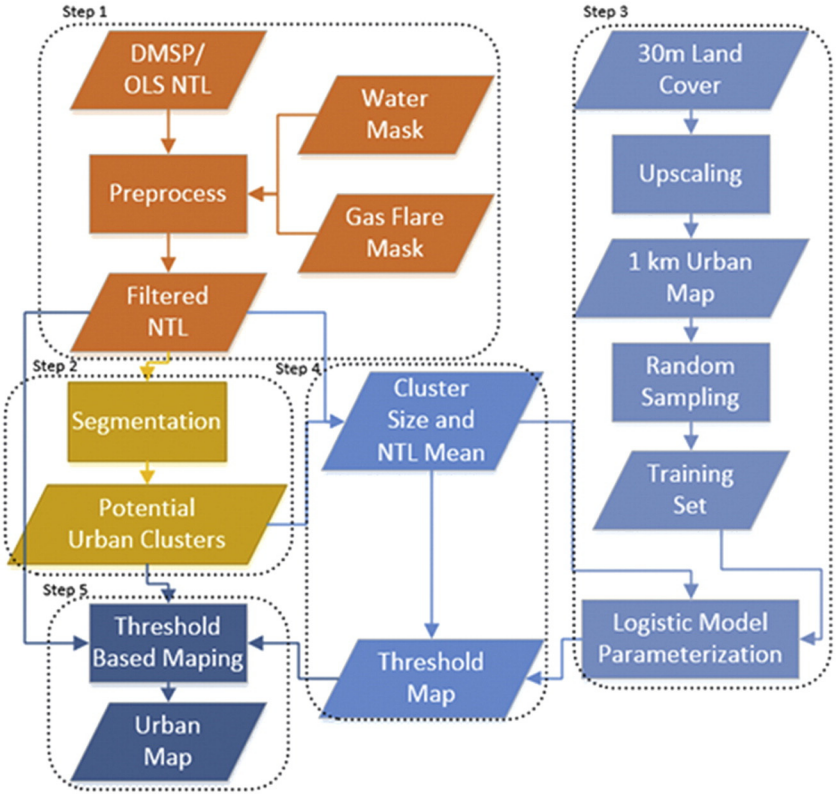

Fig. 1. Flowchart of the cluster-based method.

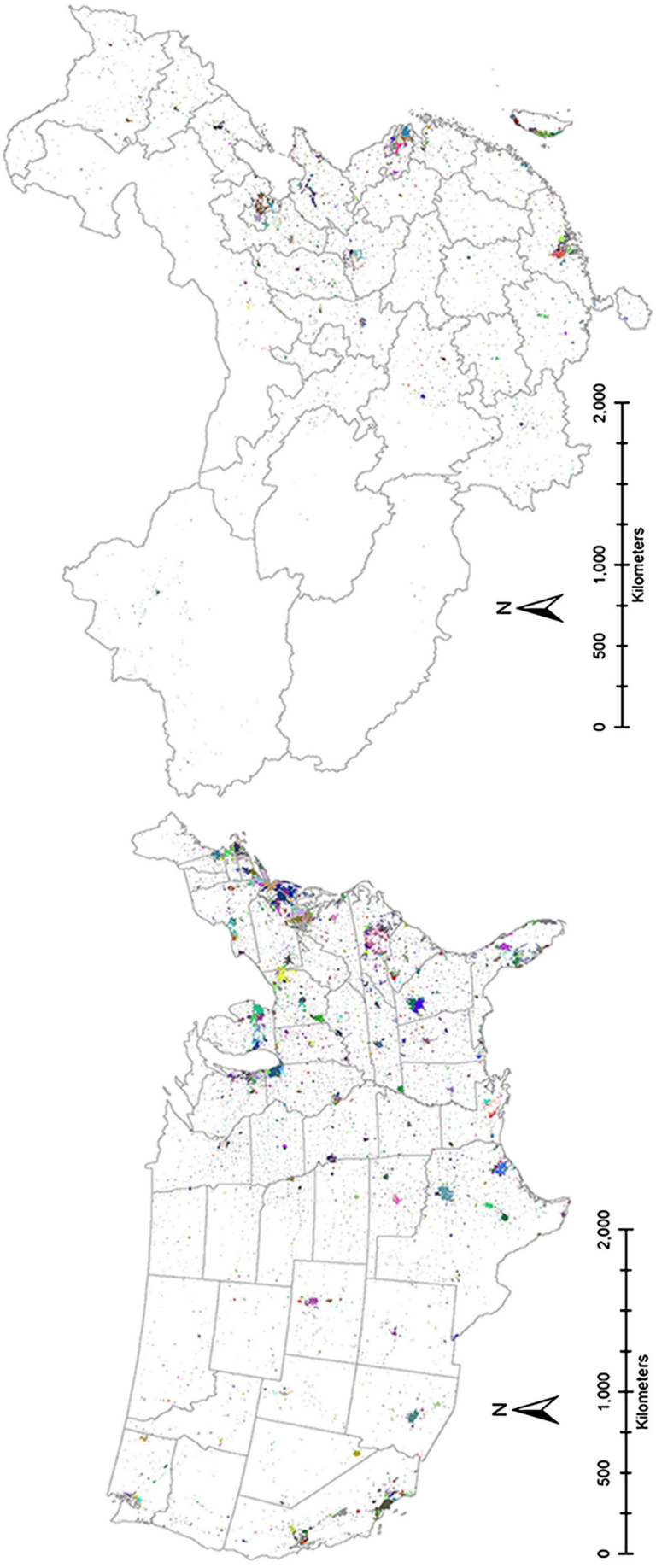

Fig. 2. Potential urban clusters; different clusters are identified by different colors. (For interpretation of the references to colors in this figure legend, the reader is referred to the web version of this article.)

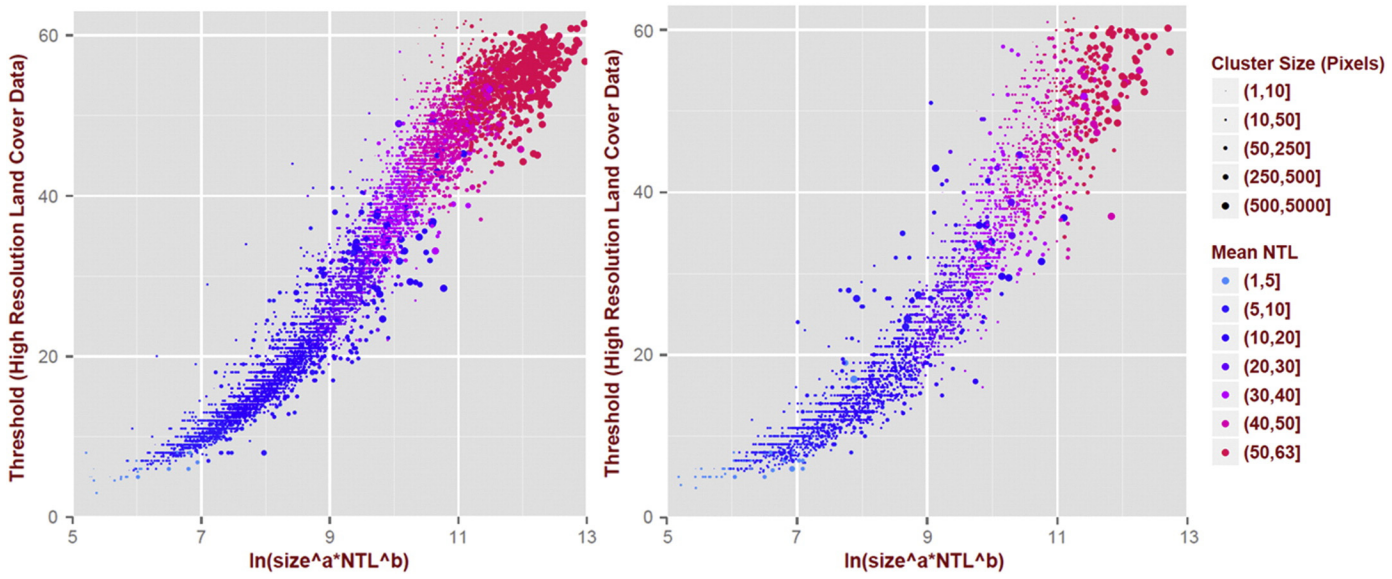

Fig. 3. Relationship between optimal thresholds derived from high resolution data and cluster size and NTL mean value in the US (left) and China (right) with $a = 0.4$ and $b = 2$ as example parameters (Eq. (1)). The cluster size is measured as number of 1-km pixels.

generally for object-based classification in remote sensing (Blaschke, 2010; Definiens, 2009; Dey et al., 2010; Hay et al., 2003; Muñoz et al., 2003; Woodcock and Harward, 1992; Zhou & Wang, 2008). For example, Woodcock and Harward (1992) developed a multiple-pass algorithm to extract forest information from multi-spectral Landsat Thematic Mapper data. Zhou and Wang (2008) improved this algorithm with additional shape information for high resolution ISA mapping. Most of these segmentation algorithms perform well driven by multiple bands of spectral information from high spatial resolution remote sensing data. Muñoz et al. (2003) presented a summary of segmentation algorithms and argued that there was no perfect segmentation algorithm for various applications. The algorithm developed by Parvati et al. (2008) is based on gray-scale morphology, which is appropriate for the single-band DMSP/OLS NTL data, similar to topographic features, and can be easily applied to large datasets. In this algorithm, a robust and flexible Marker-controlled Watershed Segmentation is used for image segmentation after morphological processes. The segmentation algorithm includes four major steps. First, a two-dimensional map of gradient magnitude was built on the filtered NTL DN data. Second, foreground markers, which are connected blobs of pixels within each of the objects, were obtained using morphological reconstruction. Third, the background markers, the dark pixels belonging to the background, were created through distance transform and watershed transform of the modified image. Finally, the watershed-based segmentation method, in which the boundaries of the objects are expressed as ridges, was applied on the processed gradient image with foreground and background markers by region growing to get the final segmented map of potential urban clusters (Fig. 2). Not all pixels in each potential cluster are urban, and the percentage of actual urban area in a potential cluster varies with clusters. Urban extent could be 0% in the cases where all pixels in a cluster fall below the threshold value. Urban pixels will be classified in each potential cluster based on the optimal threshold, which will be estimated using a logistic model in the next step. The total number of potential urban clusters in the US is more than twice that in China.

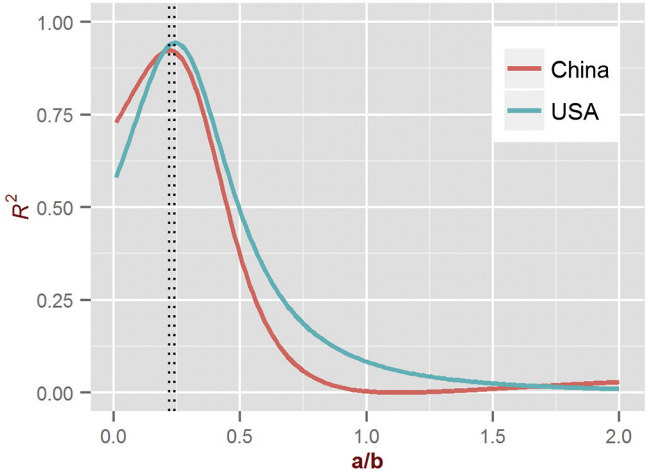

Fig. 4. Parameter selection for the US and China for the logistic model.

3.3. Logistic model

As discussed in the Introduction, urban extent is prone to being overestimated because the OLS-derived light features are substantially larger than the lighting sources on the ground due to several contributing factors. The threshold to delineate urban extent varies across countries, regions, and urban clusters. The exaggeration in the OLS NTL data is generally proportional to the lighting magnitude and cluster size. However, the relative contribution from each is unknown. Therefore, we propose an index x, which combines effects from both mean lighting magnitude and cluster size, where the contribution of each of these effects is variable and yet to be determined. We will then evaluate the relationship of this index to the optimal threshold value. Moreover, the index is transformed through a natural log to approach a normal distribution and facilitate its use over a range of magnitude values.

$$x = \ln\left(S^a NTL_{mean}{}^b\right) \quad (1)$$

where S is the cluster size, NTL_{mean} is the mean NTL DN in each cluster, and a and b are coefficients to be estimated below.

We then examined the relationship between the index x and the optimal threshold. For each cluster, the value of x is directly obtained by computing the cluster size and the mean NTL DN magnitude. The optimal threshold in each cluster was indirectly estimated by comparing

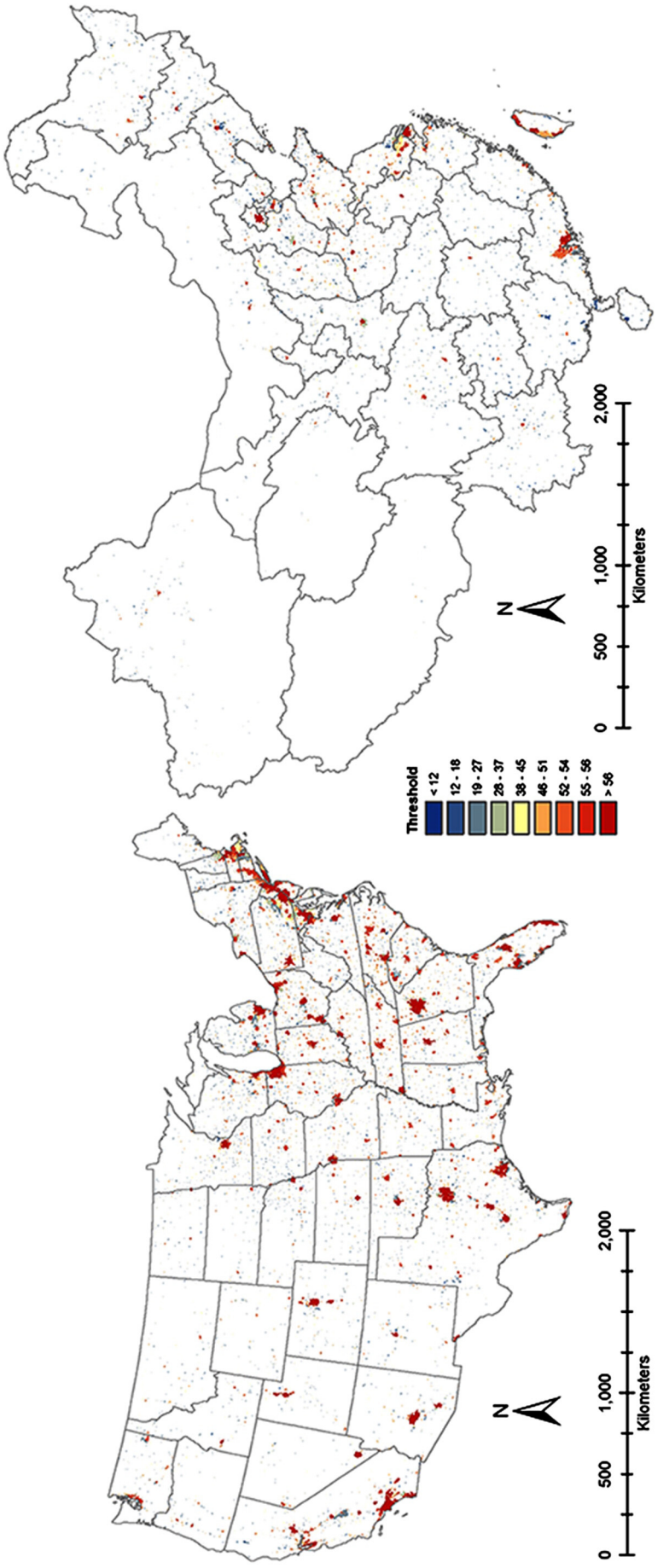

Fig. 5. Threshold maps built using the logistic model.

the 1 km binary reference urban maps with the filtered NTL data, including two steps. First, the fraction of urban area within the cluster was calculated from the 1 km binary reference map. Second, because a given threshold will lead to a unique estimated fraction of urban area for each cluster using the NTL data, the optimal threshold can be chosen as the one that can give the same fraction of urban area in the NTL cluster as that from the 1 km reference map. The so-derived optimal threshold shows a strong functional relationship with the index x. Fig. 3 shows this relationship using the values $a = 0.4$ and $b = 2$. The optimal values for these parameters will be determined in the next step.

Fig. 3 indicates a nonlinear, slightly S-shaped, relationship between the optimal threshold and index x. Motivated by this finding, we built a logistic model of the relationship between the index x and the optimal threshold (Fig. 3, Eq. (2)) as follows:

$$NTL_{\text{thld}} = \frac{1}{1 + e^{-\beta(x - x_{mean})}}(NTL_{\max} - NTL_{\min}) + NTL_{\min} \quad (2)$$

where NTL_{thld} is the optimal threshold to delineate the urban area in the potential cluster, and x_{mean} is the mean value of x, and it is calculated as the mean of all x values. $NTL_{\min}$ and $NTL_{\max}$ are minimum and maximum NTL DN in the study area, and β is the coefficient for the logistic model.

In Eq. (2), a, b, and β are unknown. To reduce the number of parameters, the index x is modified as x' (Eq. (3)).

$$x' = \ln\left(S^{\frac{a}{b}} \cdot NTL_{mean}\right). \quad (3)$$

Now, β and b can be estimated jointly in the logistic model as a single coefficient β' as shown in Eq. (4).

$$NTL_{thld} = \frac{1}{1 + e^{-\beta'(x' - x'_{mean})}}(NTL_{\max} - NTL_{\min}) + NTL_{\min}. \quad (4)$$

Eq. (4) can be converted to a linear form, and the combined coefficient β' can be calculated using ordinary least squares (OLS) regression. x'_{mean} is the mean value of x'.

$$\ln\left(\frac{NTL_{\max} - NTL_{\min}}{NTL_{thld} - NTL_{\min}} - 1\right) = -\beta' x'. \quad (5)$$

The coefficient a/b indicates the relative weight of mean NTL magnitude and cluster size contributing to the index x. The coefficient β' indicates how quickly the optimal threshold changes with changing index x.

We then examined the relationship between R^2 in Eq. (5) based on OLS regression and the coefficient a/b (Fig. 4) by using a training data subsample, which is a randomly selected half of the potential urban clusters in the US and China. The other half of the potential urban clusters will be used as a validation dataset to evaluate the thresholds derived from the proposed method. The regional value of the coefficient a/b with the highest R^2 will be used in the calculation of the index x' (Eq. (3)) in each potential urban cluster. We then calculated the coefficient β' in Eq. (4) using the same training data subsample.

3.4. Threshold estimation

The coefficients a/b and β' estimated using the training dataset are 0.24 and 0.84 in the US, and 0.22 and 0.88 in China, respectively. The difference in a/b between these two regions indicates that cluster size is a slightly more important determinant of the optimal threshold in the US as compared to China. The difference in β' between these two regions shows that the optimal threshold changes with changing index x slightly more quickly in China as compared to the US. Using the estimated coefficients a/b, we calculated x' using the cluster size and NTL DN magnitude (Eq. (3)), and then together with β', we estimated the optimal threshold in each cluster to delineate the urban extent (Eq. (4)) for all clusters in the US and China.

3.5. Urban extent mapping

With the optimal threshold estimated for each cluster, the pixels in each cluster with NTL DN larger than the optimal threshold were classified as urban and all others as nonurban. The final urban map products in the US and China were then generated.

4. Results and discussion

4.1. Threshold

The optimal thresholds tend to be larger in larger clusters with higher NTL DN in both countries (Fig. 5), which is an expected finding

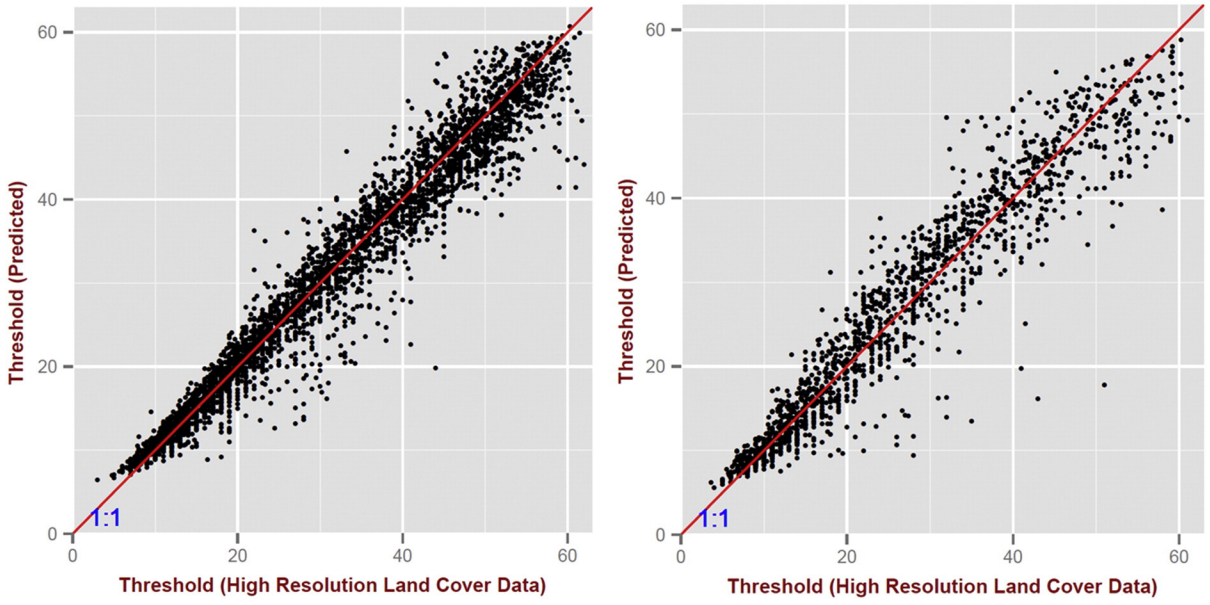

Fig. 6. Comparison of thresholds from the logistic model and high-resolution land cover data in the US (left) and China (right) for the validation data subsample.

(Fig. 3, Eq. (2)). In large urban clusters such as Boston and Beijing, the optimal thresholds reach as high as 60 while they are as low as 20 in small urban clusters. Our technique of determining optimal thresholds based on size and NTL DN magnitude can help to reduce under- and over-estimation, which has been a major issue with the use of a single threshold in regional or national studies.

To evaluate the performance of the logistic model, we use the validation subsample of potential urban clusters. For these clusters, the optimal thresholds estimated using the logistic model were compared with those derived directly from the 1 km reference binary urban maps. R^2 and Root Mean Square Error (RMSE) between the two estimates are shown in Fig. 6. It was found that the logistic model performs well in deriving the threshold in both the US (R^2: 0.96, RMSE: 3.1) and China (R^2: 0.91, RMSE: 4.2).

A sensitivity analysis was performed for a/b and β' to examine how changes in these parameters impact optimal thresholds (Fig. 7). Both parameters can have important impacts on optimal thresholds in both countries when they deviate from their optimum values. These two parameters show similar patterns of sensitivity to optimal thresholds in both the US and China. The impact of the parameter a/b directly shifts all optimal thresholds up or down from the 1:1 line. Optimal thresholds are underestimated with lower a/b while they are overestimated with higher a/b. The impact of the parameter β' rotates optimal thresholds at a certain point on the 1:1 line. Higher β' leads to underestimated optimal thresholds in the low value range and overestimated optimal threshold in the high value range. Lower β' gives the opposite effect. It is important to note that the derived optimal thresholds are not highly sensitive to the parameter choices when these parameters are sufficiently close, e.g. 20%, to the optimal values. In all cases in the figure, the reduction of R^2 is negligible (<0.01). Regarding RMSE, it increases by about 0.9 and 0.6, respectively, in the US and China when a/b deviates from the optimal values by 20%. RMSE increases by about 0.6 and

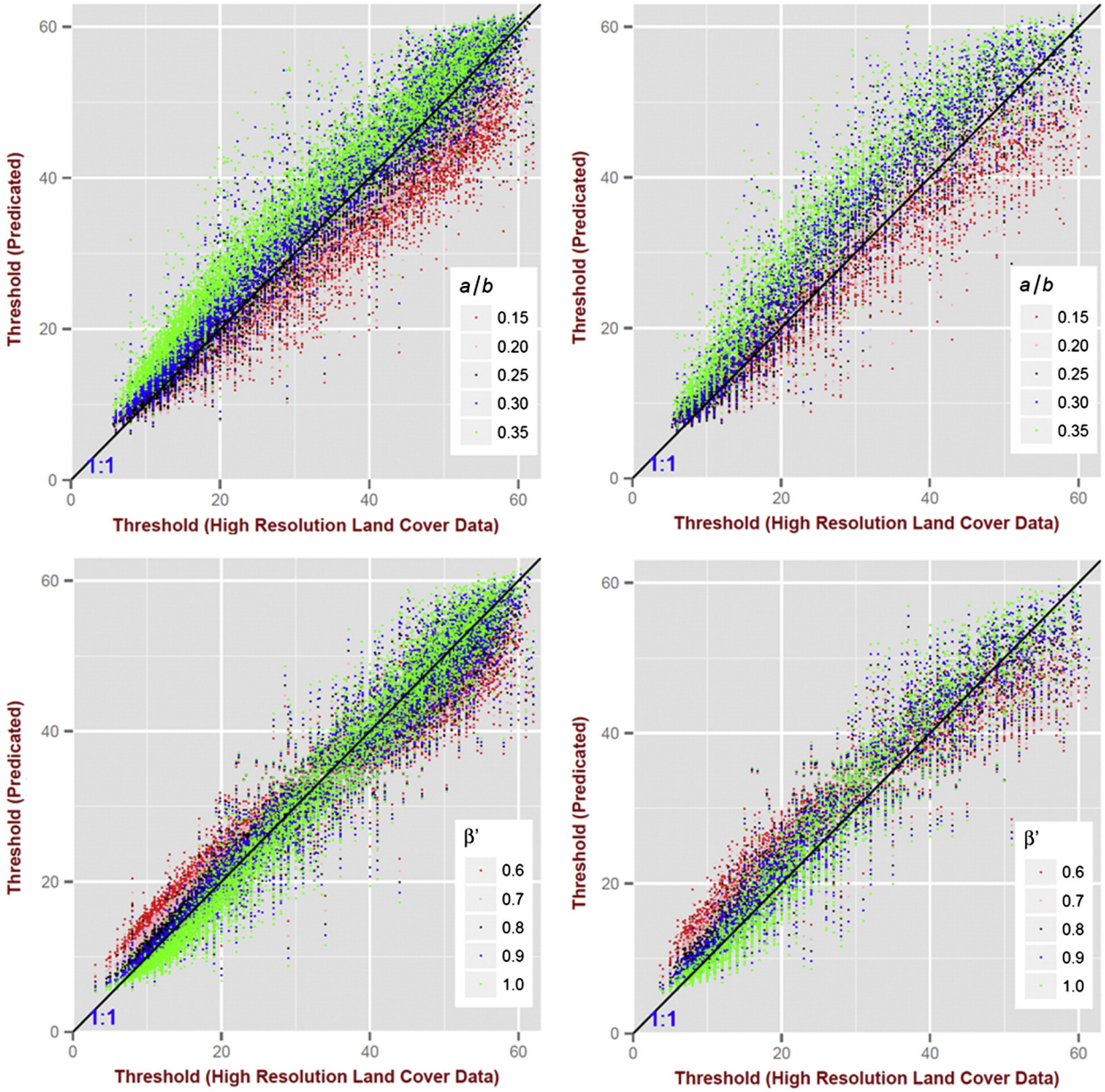

Fig. 7. Sensitivity of optimal thresholds to the optimal parameters a/b (top) and β ' (bottom) in the US (left) and China (right) for the validation data subsample.

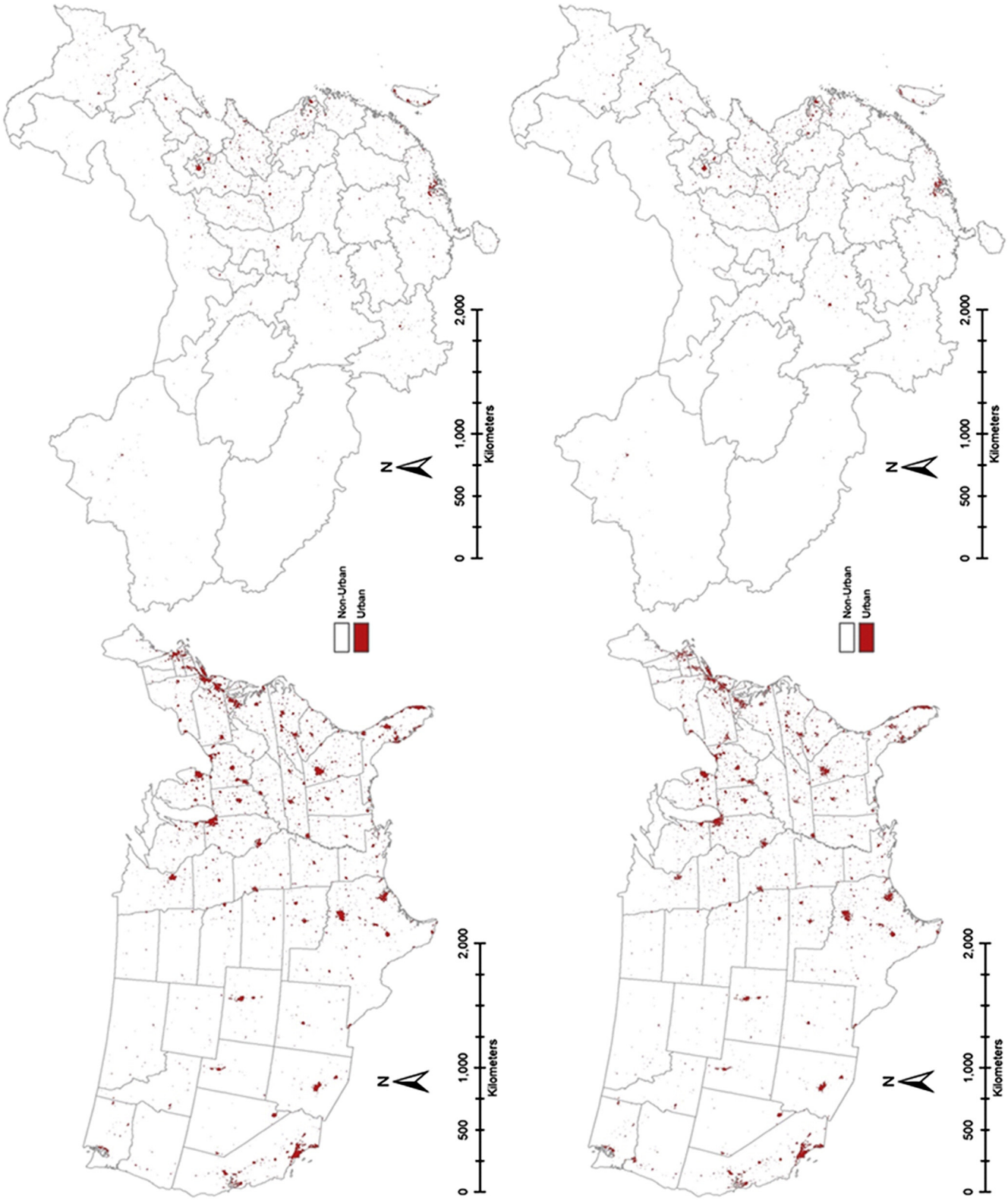

Fig. 8. Urban extent mapped from NTL (top) and high-resolution land cover data (bottom).

0.4, respectively, in the US and China when β' deviates from the optimal values by 20%. This lack of sensitivity near the optimal value can also be seen in Fig. 4.

4.2. Urban extent

With optimal thresholds calculated from the logistic model using the training data subsample, we delineated urban extent in the US and China using the complete filtered NTL data and potential urban cluster maps for each region (Fig. 8, top). The result shows the spatial patterns of urban area in the US in 2006 and China in 2005. The total urban areas are 160,000 km^2 and 37,000 km^2 in the US in 2006 and in China in 2005, respectively. Urban area in the US is more heterogeneous with centers spotted across the country, whereas in China, urban area tends to aggregate in eastern coastal regions. Urbanized area occupies about 2% of the total area in the US. The urbanization extent in terms of urban area percentage varies from lower than 0.5% in states in the US census region of Mountain and West North Central, such as Montana, Wyoming, South Dakota and North Dakota, to higher than 10% in the Eastern Coastal states such as New Jersey, Massachusetts, Connecticut, and Maryland. The urbanized area occupies less than 1% of total land area in China, and varies in extent from lower than 0.1% in some western provinces such as Xizang, Qinghai, and Xinjiang, to about 2% in the eastern coastal provinces such as Jiangshu and Guangdong. It is as high as 10% in municipalities such as Beijing and Shanghai.

A visual comparison with high-resolution urban area maps (Fig. 8, bottom) indicates that our method can delineate most of the large urban centers in both the US and China, and the extent of the urban areas from NTL and high-resolution land cover data matches relatively well (Fig. 8). We also selected four cities of different sizes in the US and China to evaluate urban extent mapped from NTL data (Fig. 9). A visual comparison shows that our proposed method can map the urban extent relatively well for example cities. The mapped urban extent tends to be less fragmented compared to those from high-resolution land cover data due to the lower spatial resolution of NTL data. Both representations of urban area have their strengths in applications. Information from high resolution data with fragmented representation of urban area is more helpful in urban studies at local scales while the information from our data, which is closer to a real world concept of a city in terms of urban area, is useful in large scale urbanization studies.

Using NTL to map urban extent can capture areas such as airports that are not classified as urban in some land-cover data, which defines

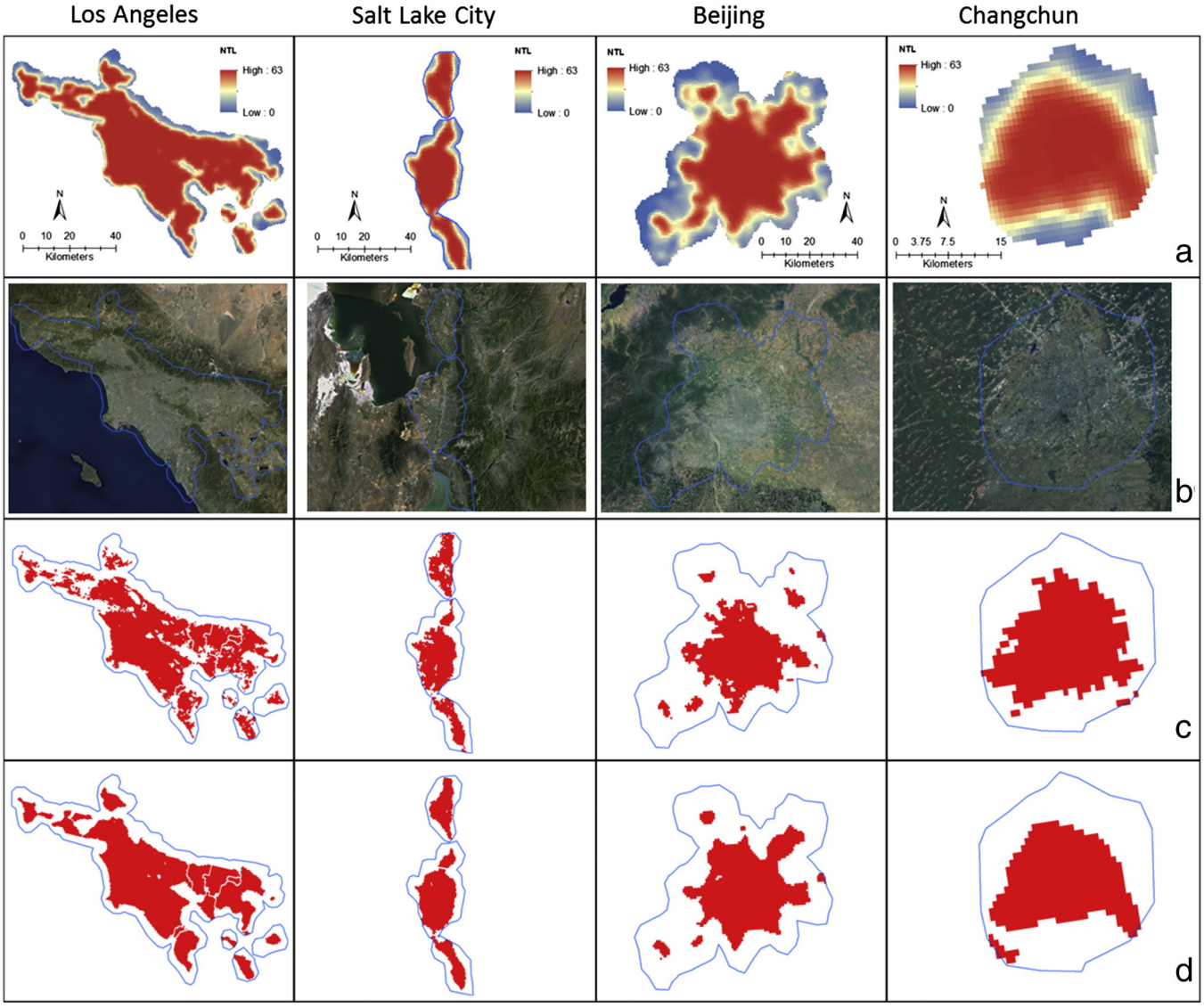

Fig. 9. Evaluation of mapped urban extent in four selected cities in the US and China: (a) DMSP NTL data, (b) Bing imageries, (c) urban extent from high-resolution land cover data, and (d) urban extent mapped from NTL data. The boundaries in blue on panels (b), (c), and (d) were derived from NTL data in panel (a). (For interpretation of the references to colors in this figure legend, the reader is referred to the web version of this article.)

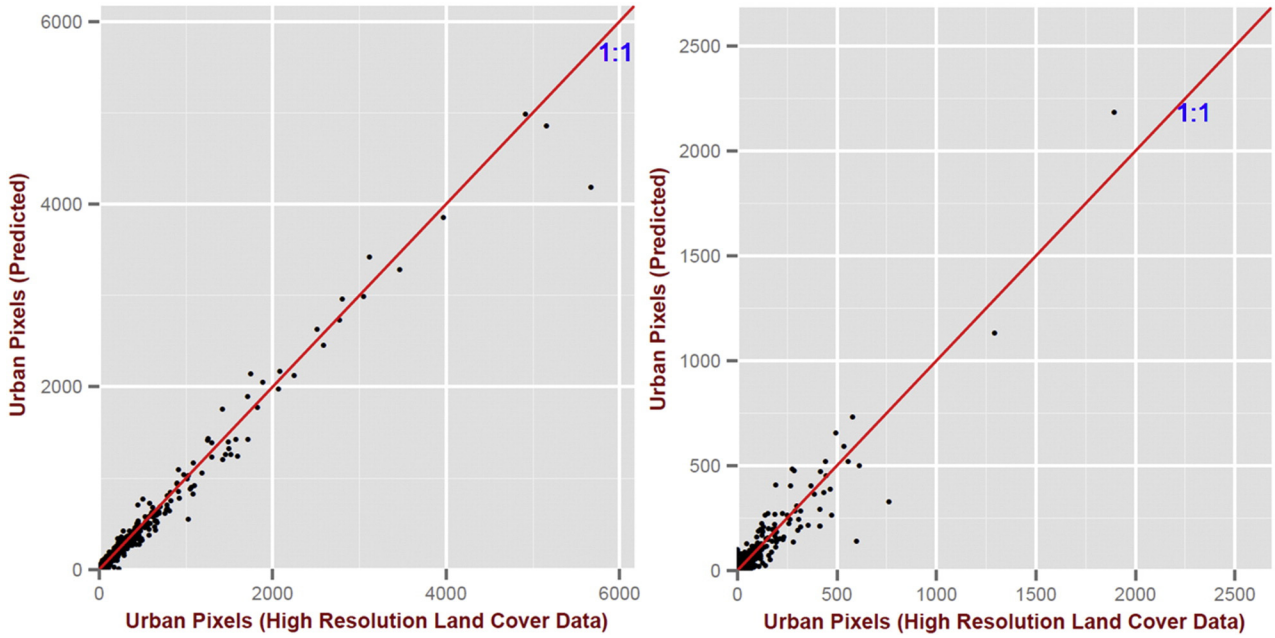

Fig. 10. Comparison of urban extent (in number of 1-km urban pixels) from the cluster-based method and 1 km reference urban map at the cluster level in the US (left) and China (right).

urban area differently. The feature to the north west of the Beijing city center in our NTL-based urban area map is the Beijing airport area, which is not classified as urban in the reference urban dataset (Fig. 9). Note that airport areas retrieved from NTL data were considered overestimates in a previous study (Sutton, Elvidge, & Obremski, 2003).

4.3. Validation of urban extent

In addition to the visual comparison, we compared the derived urban extent using this cluster-based method to those from the 1 km reference urban map at the cluster and regional levels (Figs. 10 and 11). For the US, our method performs well at the cluster level (R^2: 0.98, RMSE: 27) and the state level (R^2: 0.98, RMSE: 697 1-km pixels) (Figs. 10 and 11, left). Although it is not as good as the performance in the US, this method still performs reasonably well in China at both the cluster level (R^2: 0.85, RMSE: 24) and the province level (R^2: 0.86, RMSE: 531 1-km pixels) (Figs. 10 and 11, right). Considering the purpose of this product for large-scale urban area mapping and monitoring and also limited by the spatial resolution of NTL data, accuracy at the pixel level is also good (USA: overall accuracy of 91%, Kappa Coefficient of 0.69, producer accuracy of 76%, and user accuracy of 73%; China: overall accuracy of 93%,

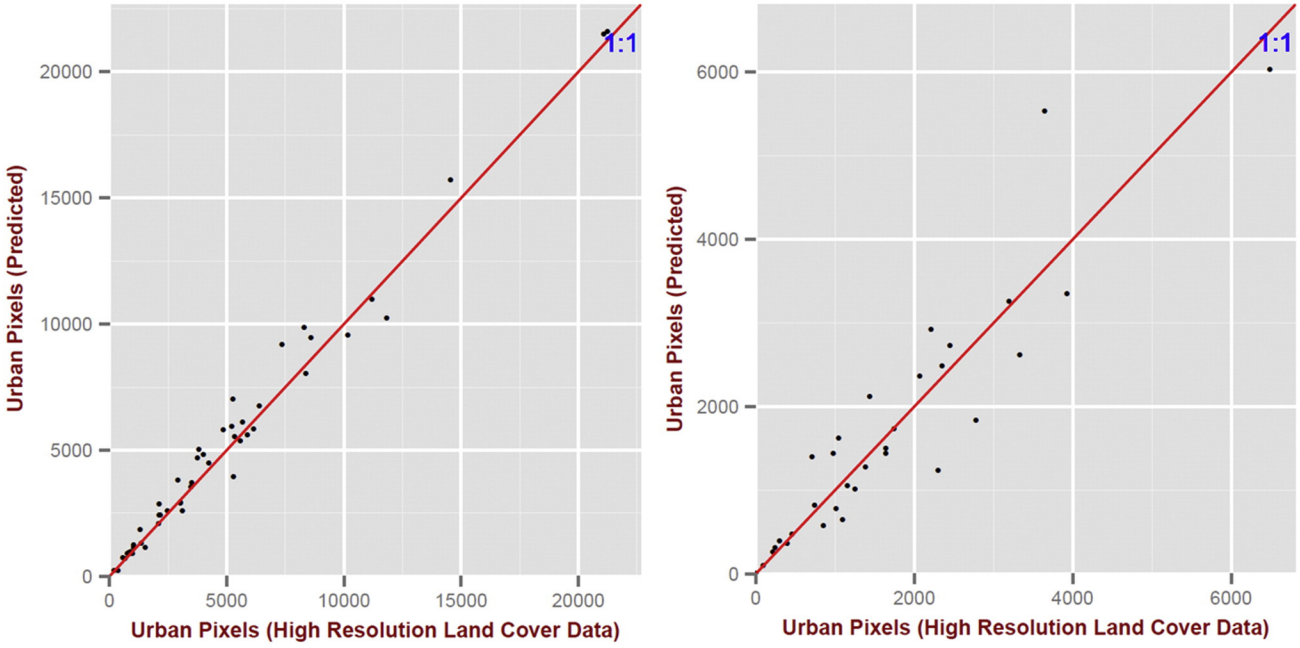

Fig. 11. Comparison of urban extent (in number of 1-km urban pixels) from the cluster-based method and 1 km reference urban map at the state/province level in the US (left) and China (right).

Kappa Coefficient of 0.54, producer accuracy of 65%, and user accuracy of 52%) though it is not as good as urban mapping using high spatial resolution data such as IKONOS. It was also found that the misclassification generally occurs on boundaries of urban clusters and green space within urban clusters.

5. Conclusions

In this study, we developed a cluster-based method to estimate optimal thresholds and delineate the urban extent from DMSP/OLS NTL data. In this method the optimal threshold for each potential urban cluster is estimated based on urban cluster size and overall nightlight magnitude in the cluster using a logistic model, resulting in thresholds specific to each urban cluster. The derived optimal thresholds are not highly sensitive to the parameter choices in the logistic model when these parameters are sufficiently close to the optimal values. The similarity of the parameters in both countries and their insensitivity in the logistic model demonstrate the potential applicability of this method more generally for global urbanization and dynamic mapping.

The total urban areas derived from the DMSP/OLS NTL data are 160,000 km^2 and 37,000 km^2 in the US in 2006 and in China in 2005, respectively. The urbanized area occupies about 2% of the total area varying from lower than 0.5% in states in the US census region of Mountain and West North Central to higher than 10% in the Eastern Coastal states. The urbanized area occupies less than 1% in China, ranging from lower than 0.1% in some western provinces to about 2% in the Eastern coastal provinces, with some municipalities as high as 10%.

Our evaluation of estimated optimal thresholds and mapped urban extent at the cluster and regional levels in the US and China confirms its utility in mapping global urban area and its dynamics using the DMSP/OLS NTL data in a timely, cost-effective way. We conclude that cluster-based optimal thresholds can map urban extent more accurately compared to global threshold techniques. Even compared to previous regional level threshold techniques, our method successfully tackles the issue of under- and over-estimation. There is no need to manually define boundaries such as economic regions with urban clusters of different sizes and NTL magnitudes, to calculate the optimal thresholds (Liu et al., 2012). Moreover, in previous regionally-based methods, some urban centers may cover multiple regions, which will result in the use of different optimal thresholds for the same urban cluster. Finally, the simplicity of this method makes it promising for rapidly monitoring urban areas globally and regionally using the DMSP/OLS NTL data.

The results presented here offer several possible avenues for future research. DMSP/OLS blooming within the clusters because of spatial resolution, overlap of adjacent pixels, and geo-location errors, are not addressed in this method (Elvidge, Baugh, Kihn, Kroehl, & Davis, 1997). Although the 1-km urban data can be upscaled to a fraction map at coarse resolution for applications such as earth-system modeling at global or regional levels (Jacobson & Ten Hoeve, 2012), the urban percentage within a pixel at 1-km cannot be derived using our proposed method. In order to address these issues, supplementary data such as MODIS NDVI may be helpful to improve urban extent mapping from the DMSP/OLS NTL data. In addition, NTL data with finer spatial resolution and higher quantization levels are becoming available, such as those from the Visible Infrared Imager Radiometer Suite (VIIRS) (Elvidge, Erwin, et al., 2009), which will potentially help to build better global urban maps using our proposed method. Given the limited temporal coverage of most global products, the method developed in this work provides a tool potentially useful to map global urban dynamics for over two decades by using DMSP/OLS NTL data. However, as the DMSP/OLS NTL products are not radiometrically calibrated, additional effort, such as inter-calibration of multiple-year NTL data, will be necessary to build a consistent NTL dataset.

Acknowledgment

We acknowledge funding support from the NASA ROSES Land-Cover/Land-Use Change Program (NNH11ZDA001N-LCLUC) with additional support for Steven J. Smith from the Global Technology Strategy Project. We would like to thank Dr. Benjamin Bond-Lamberty and the anonymous reviewers for their constructive comments and suggestions, and the many colleagues and organizations that shared the data used in this project.

References

Amaral, S., Câmara, G., Monteiro, A. M. V., Quintanilha, J. A., & Elvidge, C. D. (2005). Estimating population and energy consumption in Brazilian Amazonia using DMSP night-time satellite data. *Computers, Environment and Urban Systems*, *29*, 179–195.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*, 2–16.

Brabec, E. (2002). Impervious surfaces and water quality: A review of current literature and its implications for watershed planning. *Journal of Planning Literature*, *16*, 499–514.

Cao, X., Chen, J., Imura, H., & Higashi, O. (2009). A SVM-based method to extract urban areas from DMSP-OLS and SPOT VGT data. *Remote Sensing of Environment*, *113*, 2205–2209.

CIESIN (2011). *Global Rural–Urban Mapping Project, Version 1 (GRUMPv1): Urban Extents Grid.*

Definiens, A. G. (2009). Definiens Developer User Guide. Munich, Germany: Definiens AG.

Dey, V., Zhang, Y., & Zhong, M. (2010). *A review on image segmentation techniques with remote sensing perspective.* In, Proceedings of the International Society for Photogrammetry and Remote Sensing Symposium (ISPRS10) (pp. 5–7).

Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., & Worley, B. A. (2000). LandScan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing*, *66*, 849–857.

Doll, C. N. H., Muller, J. P., & Elvidge, C. D. (2000). Night-time imagery as a tool for global mapping of socioeconomic parameters and greenhouse gas emissions. *AMBIO: A Journal of the Human Environment*, *29*, 157–162.

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., & Davis, E. R. (1997). Mapping city lights with nighttime data from the DMSP Operational Linescan System. *Photogrammetric Engineering and Remote Sensing*, *63*, 727–734.

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., Davis, E. R., & Davis, C. W. (1997). Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing*, *18*, 1373–1379.

Elvidge, C. D., Erwin, E. H., Baugh, K. E., Ziskin, D., Tuttle, B. T., Ghosh, T., et al. (2009). Overview of DMSP nightime lights and future possibilities. *Urban remote sensing event, 2009 joint* (pp. 1–5). : IEEE.

Elvidge, C. D., Safran, J., Tuttle, B., Sutton, P., Cinzano, P., Pettit, D., et al. (2007). Potential for global mapping of development via a nightsat mission. *GeoJournal*, *69*, 45–53.

Elvidge, C. D., Sutton, P. C., et al. (2009). Global urban mapping based on nighttime lights. In P. Gamba, & M. Herold (Eds.), *Global mapping of human settlement* (pp. 129–144). Boca Raton, Florida: Taylor and Francis Group.

Elvidge, C. D., Tuttle, B. T., Sutton, P. C., Baugh, K. E., Howard, A. T., Milesi, C., et al. (2007). Global distribution and density of constructed impervious surfaces. *Sensors*, *7*, 1962–1979.

Elvidge, C., Ziskin, D., Baugh, K., Tuttle, B., Ghosh, T., Pack, D., et al. (2009). A fifteen year record of global natural gas flaring derived from satellite data. *Energies*, *2*, 595–622.

ESA (2013). *GlobCover land cover maps.*

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., et al. (2005). Global consequences of land use. *Science*, *309*, 570–574.

Frolking, S., Milliman, T., Seto, K. C., & Friedl, M. A. (2013). A global fingerprint of macro-scale changes in urban structure from 1999 to 2009. *Environmental Research Letters*, *8*, 024004.

Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., et al. (2011). Completion of the 2006 national land cover database for the conterminous United States. *PE&RS*, *77*(9), 858–864 (In: D).

Hay, G. J., Blaschke, T., Marceau, D. J., & Bouchard, A. (2003). A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS Journal of Photogrammetry and Remote Sensing*, *57*, 327–345.

He, C., Shi, P., Li, J., Chen, J., Pan, Y., Li, J., et al. (2006). Restoring urbanization process in China in the 1990s by using non-radiance-calibrated DMSP/OLS nighttime light imagery and statistical data. *Chinese Science Bulletin*, *51*, 1614–1620.

Henderson, M., Yeh, E. T., Gong, P., Elvidge, C., & Baugh, K. (2003). Validation of urban boundaries derived from global night-time satellite imagery. *International Journal of Remote Sensing*, *24*, 595–609.

Homer, C., Huang, C., Yang, L., Wylie, B. K., & Coan, M. (2004). *Development of a 2001 national land-cover database for the United States.*

Imhoff, M. L., Bounoua, L., DeFries, R., Lawrence, W. T., Stutzer, D., Tucker, C. J., et al. (2004). The consequences of urban land transformation on net primary productivity in the United States. *Remote Sensing of Environment*, *89*, 434–443.

Imhoff, M. L., Lawrence, W. T., Stutzer, D. C., & Elvidge, C. D. (1997). A technique for using composite DMSP/OLS "City Lights" satellite data to map urban area. *Remote Sensing of Environment*, *61*, 361–370.

Jacobson, M. Z., & Ten Hoeve, J. E. (2012). Effects of urban surfaces and white roofs on global and regional climate. *Journal of climate*, *25*, 1028–1044.

Kasimu, A., Tateishi, R., & Hoan, N. (2009). Global urban characterization using population density, DMSP and MODIS data. *2009 Urban Remote Sensing Joint Event*.
Liu, Z., He, C., Zhang, Q., Huang, Q., & Yang, Y. (2012). Extracting the dynamics of urban expansion in China using DMSP-OLS nighttime light data from 1992 to 2008. *Landscape and Urban Planning*, *106*, 62–72.
Liu, J., Zhang, Z., Xu, X., Kuang, W., Zhou, W., Zhang, S., et al. (2010). Spatial patterns and driving forces of land use change in China during the early 21st century. *Journal of Geographical Sciences*, *20*, 483–494.
Loveland, T., Reed, B., Brown, J., Ohlen, D., Zhu, Z., Yang, L., et al. (2000). Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, *21*, 1303–1330.
Lu, D., Tian, H., Zhou, G., & Ge, H. (2008). Regional mapping of human settlements in southeastern China with multisensor remotely sensed data. *Remote Sensing of Environment*, *112*, 3668–3679.
McKinney, M. L. (2008). Effects of urbanization on species richness: A review of plants and animals. *Urban Ecosystems*, *11*, 161–176.
Muñoz, X., Freixenet, J., Cufi, X., & Marfi, J. (2003). Strategies for image segmentation combining region and boundary information. *Pattern Recognition Letters*, *24*, 375–392.
Nizeyimana, E. L., Petersen, G., Imhoff, M., Sinclair, H., Waltman, S., Reed-Margetan, D., et al. (2001). Assessing the impact of land conversion to urban use on soils with different productivity levels in the USA. *Soil Science Society of America Journal*, *65*, 391–402.
Oda, T., & Maksyutov, S. (2011). A very high-resolution (1 km × 1 km) global fossil fuel CO_2 emission inventory derived using a point source database and satellite observations of nighttime lights. *Atmospheric Chemistry and Physics*, *11*, 543–556.
Owen, T. (1998). Using DMSP-OLS light frequency data to categorize urban environments associated with US climate observing stations. *International Journal of Remote Sensing*, *19*, 3451–3456.
Parshall, L., Gurney, K., Hammer, S. A., Mendoza, D., Zhou, Y., & Geethakumar, S. (2010). Modeling energy consumption and CO_2 emissions at the urban scale: Methodological challenges and insights from the United States. *Energy Policy*, *38*, 4765–4782.
Schneider, A., Friedl, M. A., & Potere, D. (2010). Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions'. *Remote Sensing of Environment*, *114*, 1733–1746.
Parvati, K., Prakasa Rao, B. S., Mariya Das, M., et al. (2008). Image Segmentation Using Gray-Scale Morphology and Marker-Controlled Watershed Transformation. Discrete Dynamics in Nature and Society.
Schneider, A., & Woodcock, C. E. (2008). Compact, dispersed, fragmented, extensive? A comparison of urban growth in twenty-five global cities using remotely sensed data, pattern metrics and census information. *Urban Studies*, *45*, 659–692.
Shepherd, J. M. (2005). A review of current investigations of urban-induced rainfall and recommendations for the future. *Earth Interactions*, *9*, 1–27.
Small, C., Pozzi, F., & Elvidge, C. D. (2005). Spatial analysis of global urban extent from DMSP-OLS night lights. *Remote Sensing of Environment*, *96*, 277–291.
Sutton, P. C. (2003). A scale-adjusted measure of "urban sprawl" using nighttime satellite imagery. *Remote Sensing of Environment*, *86*, 353–369.
Sutton, P. C., Cova, T. J., & Elvidge, C. D. (2006). Mapping "exurbia" in the conterminous United States using nighttime satellite imagery. *Geocarto International*, *21*, 39–45.
Sutton, P. C., Elvidge, C., & Obremski, T. (2003). Building and evaluating models to estimate ambient population density. *Photogrammetric Engineering & Remote Sensing*, *69*, 545–553.
Woodcock, C., & Harward, V. (1992). Nested-hierarchical scene models and image segmentation. *International Journal of Remote Sensing*, *13*, 3167–3187.
Zhang, Q., Schaaf, C., & Seto, K. C. (2013). The Vegetation Adjusted NTL Urban Index: A new approach to reduce saturation and increase variation in nighttime luminosity. *Remote Sensing of Environment*, *129*, 32–41.
Zhang, Q., & Seto, K. C. (2011). Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. *Remote Sensing of Environment*, *115*, 2320–2329.
Zhou, Y., & Wang, Y. (2007). An assessment of impervious surface areas in Rhode Island. *Northeastern Naturalist*, *14*, 643–650.
Zhou, Y., & Wang, Y. (2008). Extraction of impervious surface areas from high spatial resolution imageries by multiple agent segmentation and classification. *PE&RS*, *74*, 857–868.
Zhou, Y., Wang, Y., Gold, A. J., & August, P. V. (2010). Modeling watershed rainfall–runoff relations using impervious surface-area data with high spatial resolution. *Hydrogeology Journal*, *18*, 1413–1423.
Zhou, Y., Wang, Y., Gold, A. J., August, P. V., & Boving, T. B. (2013). Assessing impact of urban impervious surface on watershed hydrology using distributed object-oriented simulation and spatial regression. *GeoJournal*. http://dx.doi.org/10.1007/s10708-10013-19506-x.
Zhou, Y., Weng, Q., Gurney, K. R., Shuai, Y., & Hu, X. (2012). Estimation of the relationship between remotely sensed anthropogenic heat discharge and building energy use. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*, 65–72.