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

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