Remote Sensing Applications: Society and Environment 10 (2018) 93–103

Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

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

Estimating urban areas: New insights from very high-resolution human settlement data☆

Pranab K. Roy Chowdhury[a,b,c,*], Budhendra L. Bhaduri[a,b], Jacob J. McKee[a,b]

[a] Urban Dynamics Institute, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[b] Geographic Information Science and Technology Group, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[c] Bredesen Center For Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN, USA

ARTICLE INFO

Keywords:
Human settlement
Built-up area estimate
Urban dynamics
High-resolution mapping

ABSTRACT

Estimation of the built-up areas is fundamental to studying urbanization and the concomitant impacts on our rapidly changing planet. Local-scale mapping of the built-up areas elucidates spatially distinct patterns of urban densification as well as peri-urban growth, where the most socially vulnerable population traditionally resides, and helps to ensure sustainable and equitable urban development. Due to the scale and spatial dependency of urban processes, most land use and land cover (LULC) data produced at national and regional scales cannot adequately capture this local variation which is readily observed in very high-resolution (≤0.5 m) remotely sensed images.

Our study investigates whether human settlement data derived from very high-resolution images provide unique understanding in the mapping of built-up areas and further the knowledge of human signatures at local levels. We selected two disparate geographies, Egypt and Taiwan, for which we analyzed four datasets representing human settlements at different spatial resolutions. Our analysis of urban morphology is based on aggregation, complexity, and contiguity of built-up areas on these settlement data and conducted at multiple spatial scales corresponding to the original resolution of the datasets. The results indicate that estimates of the total built-up area are severely misconceived, with most anomalies occurring along fringe areas. This work also illustrates the potential of high-resolution datasets to provide new insight into urban dynamics, through determining new measures of built-up area and identifying complex urban and peri-urban patterns that were previously undetected.

1. Introduction

Urban dwellers constituted 54% of global population in 2014, which is projected to reach 66% by the year 2050 (United Nations, 2014). This unprecedented growth necessitates efforts to ensure equity, shared prosperity, and welfare for all. A key step towards achieving such a noble goal is a timely and accurate mapping of built-up areas. Identification of new urban areas resulting from the peri-urban growth, as well as from infilling of existing urban areas is important as this is where the most vulnerable members of urban communities are likely to be located. Human settlement maps derived from remotely sensed images are better suited for this purpose over traditional survey or census-based methods (Schneider and Woodcock, 2008). Over past three decades such datasets have been produced at various spatial scales such as the 1 km Global Rural-Urban Mapping Project Version 1 (GRUMPv1) urban extent grids (CIESIN, 2011), 1 km Global Impervious Surface Area data from the United States Geophysical Data Center (Elvidge et al., 2007), 1 km and 500 m MODIS land cover data products (Friedl et al., 2010; Channan et al., 2014), and 250 and 100 m CORINE (Coordination of Information on The Environment) datasets by European Environment Agency (Perdigao and Annoni, 1997). These free and open access global datasets have made strong scientific contributions in a range of studies related to population distribution modeling (Linard et al., 2012), environmental impacts (Schneider et al., 2010), urban planning (Taubenbock et al., 2009), disaster management (Fleiss et al., 2011), health (Bailey et al., 2011), and energy studies (Zhang and Seto, 2011). However, the low-resolution datasets fail to identify finer details required for urban dynamics research. Thus proving insufficient

* Corresponding author at: Bredesen Center For Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN, USA.

https://doi.org/10.1016/j.rsase.2018.03.002
Received 3 October 2017; Received in revised form 7 March 2018; Accepted 8 March 2018
Available online 13 March 2018

in accounting for large parts of built-up patches in and around urban areas across the globe. Additionally, the lack of regular updates for most human settlement datasets seriously hampers the urban change detection and monitoring related work.

Increased availability of very high-resolution remotely sensed images in recent years, coupled with advancements in high-performance computing resources and efficient image processing algorithms have fostered the development of high-resolution human settlement datasets (Cheriyadat et al., 2007; Vijayaraj et al., 2007; Patlolla et al., 2012). Examples of such newly developed datasets include Global Human Settlement Layer (GHSL), Global Urban Footprint (GUF), and LandScan Settlement Layer (LandScan SL). GHSL, produced by the European Commission, is derived from cross-platform and multi-sensor image data. It is developed using a new classifying method based on symbolic machine learning (Corbane et al., 2017; Pesaresi et al., 2016b) and texture extraction method known as PANTEX (Klotz et al., 2016; Pesaresi et al., 2008). The data is available at 38.2 m (fine) and 305.8 m (aggregated) resolutions (Pesaresi et al., 2016a). Recently, the German Aerospace Center has developed a high-resolution dataset, GUF, available at an approximate spatial resolution of 12 m (fine) and 84 m (aggregated) (Esch et al., 2012). GUF has been produced using local speckle analysis on 3 m radar data from TerraSAR-X and TanDEM-X missions (Klotz et al., 2016). Both GHSL and GUF are freely available at global scale for scientific purposes from their respective producers. Furthermore, Oak Ridge National Laboratory (ORNL) has been developing a human settlement layer, LandScan SL, in support of the LandScan project and the spatial refinement of gridded population datasets. With an approximate spatial resolution of 8 m, LandScan SL is produced using supervised pattern recognition from very high-resolution (≤ 0.5 m) satellite images (Cheriyadat et al., 2007; Patlolla et al., 2012).

2. Estimation of built-up area extents

Past researchers have noted significant discord among different datasets regarding the estimation of built-up and urban areas (Schneider et al., 2003, 2010; Potere et al., 2009; Herold, 2009). For example, a study involving eight global level coarse and medium-resolution datasets revealed a range of global urban area estimates starting from 0.27×10^6 to 3.52×10^6 km^2 (Potere et al., 2009). It should be noted that the scale of representation, whether global or regional, could potentially have a bearing on the variation across estimates (Klotz et al., 2016). The long-standing issues surrounding urban area detection could, therefore, result in unwarranted discrepancies, such as failure to detect built-up areas, inaccuracies in derivative datasets, or potentially erroneous and unreliable results, and conclusions from subsequent analysis using these datasets (Fang et al., 2006). As human settlements exhibit complex form and uneven distribution over space, their precise detection is required to accurately convey the urban dynamics through derived products (Vatsavai, 2013). The precision of built-up area detection is a function of the input remote sensing image characteristics and methods of information extraction (Hay et al., 2005). Spatial resolution of input images determines how much information is there to be extracted. The advantage of high-resolution satellite images (<5 m) in capturing complex landscape patterns over medium (80–15 m) and low (>100 m) datasets have been highlighted in published literature (Peres et al., 2006; Boyd and Foody, 2011; Klotz et al., 2016). Comprehensive performance analysis of high-resolution settlement datasets over others is now needed to understand their advantages in scientific applications, especially regarding the capture and quantification of human settlements across variegated geographies.

Scientific studies addressing this issue remain comparatively scarce. Recently, 12 m GUF and GHSL datasets were analyzed alongside 500 m MODIS global urban extent data and 300 m GlobCover data over two test sites in central Europe to estimate the consistency of high-resolution maps to existing low and medium-resolution data. The study assessed how well the complex settlement patterns were captured by high-resolution datasets, and analyzed their performance over urban and rural landscapes with varying settlement size and density (Klotz et al., 2016). To our knowledge, however, no published research has evaluated the performance of a broad set of datasets over vastly different geographies, which offer distinct scene characteristics and challenges for remote sensing based data extraction. Also, no research has been published assessing the performance of the LandScan SL data in comparison with other datasets. In this work, we evaluated a set of four distinct datasets, consisting of high-resolution LandScan SL and GHSL alongside two other datasets with relatively lower spatial resolution. While the other three datasets are available as global coverage, due to continuing development, global coverage of LandScan SL has yet to be finalized. Thus the analysis was carried out over a subset of the data, representing two disparate geographies, Egypt and Taiwan, at national and sub national levels. We generated composite maps to identify inter-dataset correspondence regarding built-up area detection at multiple scales corresponding to the native spatial resolution of the datasets. Followed by a quantitative analysis of urban morphology using spatial metrics to draw insight on the congruency of these datasets in capturing the complex settlement patterns and supplement existing knowledge of global human settlement mapping initiatives.

3. Datasets used

3.1. MODIS land cover data

500 m MODIS land cover dataset (also known as MCD12Q1, MODIS LC hereafter) is developed by the National Aeronautics and Space Administration (NASA) (Friedl et al., 2010). MODIS LC is produced using a supervised decision tree classification algorithm on yearly eight day composites of MODIS Normalized BRDF-Adjusted Reflectance (NBAR) and MODIS Land Surface Temperature (LST) data (Strahler et al., 1999). The data preparation process also includes an optimization method termed “boosting” for improving classifier accuracy through systematic variation of training samples (Friedl et al., 2010). This data has been used extensively in urban land cover related studies at global scales (Schwarz et al., 2011; Jacobson and Ten Hoeve, 2012; Zhang et al., 2016; Miyazaki et al., 2016). We extracted the urban pixels, classified according to the International Geosphere-Biosphere Programme (IGBP) scheme, from the 2011 dataset. MODIS LC data may be freely downloaded from https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd12q1.

3.2. ESA - climate change initiative land cover data

The Climate Change Initiative (CCI) project at the European Space Agency has developed a global land cover dataset (hereafter CCI-LC) at 300 m spatial resolution. The datasets have been developed for three time periods of 2008–2012, 2003–2007, and 1998–2002, respectively centered on 2010, 2005, and 2000 (Herold et al., 2011). Medium Resolution Imaging Spectrometer (MERIS) full resolution datasets were used as the primary input while reduced resolution data were used in the absence of full resolution coverage. *Satellite Pour lObservation de la Terre* Vegetation (SPOT-VGT) data was used to fill in gaps in MERIS temporal coverage. These datasets aim to deliver stable time-series land cover maps, and follows the United Nations' Land Cover Classification Scheme (UN-LCCS) for classification of land cover classes (Herold et al., 2011; Kirches et al., 2014). We extracted the urban areas based on pixel values from the 2008–2012 dataset, which is freely accessible at https://www.esa-landcover-cci.org/.

3.3. Global human settlement layer

The Joint Research Center at the European Commission recently developed a human settlement dataset called the GHSL, utilizing Landsat time series data collection ranging from 1975 to Landsat-8