Mapping Urban Land Cover: A Novel Machine Learning Approach Using Landsat and Nighttime Lights

Ran Goldblatt, Michelle F. Stuhlmacher, Beth Tellman, Nicholas Clinton, Gordon Hanson, Matei Georgescu, Chuyuan Wang, Fidel Serrano-Candela, Amit K. Khandelwal, Wan-Hwa Cheng, Robert C. Balling, Jr.

Highlights:

- An efficient machine-learning approach is proposed to map built-up areas at large scales
- Our transfer-learning approach utilizes nighttime-lights data and Landsat imagery
- The approach overcomes the lack of extensive ground-truth data for urban research
- Hexagonal tessellation partition improves classification of heterogeneous land cover
- High quality maps of built-up areas are produced for 3 geographically diverse countries

Mapping Urban Land Cover:
A Novel Machine Learning Approach Using Landsat and Nighttime Lights

Abstract

Reliable large-scale representations of contemporary urban extent remain limited, hindering scientific progress across a range of disciplines aimed at helping create functional and sustainable cities. We present a novel, efficient, and low-cost machine-learning approach to map urban areas at large scales. Our methodology combines nighttime-lights data and Landsat 8 imagery using a transfer-learning approach that overcomes the lack of extensive ground-truth data. We demonstrate the effectiveness of our methodology through the development of high-quality 30m resolution maps that characterize urban areas in three diverse countries: India, Mexico, and the US. We implement our methodology in Google Earth Engine and show that it produces accurate maps of built-up land cover at high resolution over large spatial extents. Our approach highlights the usefulness of machine-learning techniques for studying the built environment, with broad implications for identification of urbanization drivers and effects on earth-system processes.

Keywords: Urbanization, built-up land cover, nighttime light, image classification

1. Introduction

Urbanization has been a fundamental trend of the past two centuries and a key force shaping almost every dimension of the modern world. In the period between 1950 and 2014, the share of the global population living in urban areas increased from 30% to 54%, and in the next few decades is projected to expand by an additional 2.5 billion urban dwellers, primarily in Asia and Africa (Seto et al., 2012; UN, 2014). Urban population growth is accompanied by a dramatic increase in the land area incorporated in cities (Georgescu et al., 2015). While urbanization in rapidly growing nations is helping lift hundreds of millions of people out of poverty, it is also creating immense societal challenges by increasing greenhouse-gas emissions, destabilizing fragile ecosystems and creating new demands on public services and infrastructure. Despite the importance of understanding the drivers of urban growth, we are still unable to quantify the magnitude and pace of urbanization in a consistent manner at high resolution and global scale. Standard empirical approaches use data from household surveys that are expensive to collect, produced infrequently, and subject to measurement problems.

The revolution in geospatial data transforms how we study cities. Since the 1970s, terrestrial Earth-observation data has been continuously collected in various spectral, spatial and temporal resolutions. As improved satellite imagery becomes available, new remote-sensing methods and machine-learning approaches have been developed to convert terrestrial Earth-observation data into meaningful information about the nature and pace of change of urban landscapes and human settlements (CIESIN, 2005; Gaughan et al., 2013; Pesaresi et al., 2016; Potere et al., 2009; Seto et al., 2011; Taubenböck et al., 2012). Existing maps of urban land show considerable disagreement on the location and extent of urbanization (Potere et al., 2009; Seto et al., 2011) and are further subject to limitations across space and time. These inconsistencies may arise in part because the delineation of urban land depends on the nature of the input data (Schneider et al., 2010), which may capture different dimensions of urbanization, such as built-up land cover or land use and population density (Bagan and Yamagata, 2014; Stevens et al., 2015; Tatem et al., 2007).

Since the early 1990's, data on nighttime lights have been amassed, primarily from sensors on board the Operational Line-scan System of the Defense Meteorological Satellite Program (DMSP-OLS). DMSP-OLS sensors capture artificial lighting, which is associated with developed land (Elvidge et al., 2014; Levin and Duke, 2012; Sutton, 2003) and can be used to infer the extent of urban areas (Bagan and Yamagata, 2015; Small and Elvidge, 2013; Zhang and Seto, 2013), as well as economic activity at the local, regional and national levels (Elvidge et al., 2014; Henderson et al., 2012; Keola et al., 2015). According to this approach, a pixel is considered urbanized if its magnitude exceeds a threshold, where the appropriate threshold may vary across countries (Small and Elvidge, 2013) and even across regions within a country (Henderson et al., 2003; Liu et al., 2016; Su et al., 2015; Wei et al., 2014; Zhou et al., 2015, 2014). Thus, inference using nighttime-light data is often inaccurate, especially in low-density urban areas (Zhang and Seto, 2013). DMSP-OLS can also exaggerate the extent of urban areas (Henderson et al., 2003; Small et al., 2005) while overlooking small or developing settlements. In addition, the extent and intensity of lit areas cannot directly delimit urban regions due to the "blooming" effect (Imhoff et al., 1997) and the "saturation" of the pixels (Hsu et al., 2015). "Blooming" refers to the identification of lit areas as consistently larger than the settlements they are associated with (Small et al., 2005); "saturation" occurs when pixels in bright areas, such as in city centers, reach the highest possible digital number (DN) value (i.e., 63) and no further details can be recognized (Hsu et al., 2015).

1.1. Detecting Urbanization Processes by Means of Machine Learning Approaches

Urban areas can be detected in satellite imagery using various machine-learning approaches (e.g., supervised, unsupervised and semi-supervised). These approaches typically rely on ground-truth data that mark urban features, either for training or for validation. Several datasets have been previously proposed to serve as ground truth for urban research. These include Landsat-based urban maps (Potere et al., 2009), census-based population databases (Stevens et al., 2015) and hand-labeled examples (Goldblatt et al., 2016), as well as data collected via crowd-source platforms, such as OpenStreetMap (OSM) (Belgiu and Drăgu

Despite significant progress in machine learning and the increasing availability of satellite data that can be used as input for classification, there remains a paucity of ground-truth datasets that have been developed to detect urban areas (Miyazaki et al., 2011). Previous studies have used ground-truth datasets that are of limited size (Goldblatt et al., 2016; Trianni et al., 2015). Transfer learning (or knowledge transfer) has emerged as a useful framework for machine learning and image classification, including in remote sensing, in order to address the scarcity of ground-truth data and to minimize the need for expensive labeling efforts. Transfer learning aims to transfer knowledge between task domains, even where the training and test data are drawn from a different feature space (Pan and Yang, 2010). Poverty prediction is an example of previous application of transfer learning in the remote-sensing field (Jean et al., 2016).

Until recently, the majority of studies that analyze urbanization have been limited in scale because of the lack of extensive high-resolution satellite data, scarcity of ground-truth data, and computational constraints. However, emerging cloud-based computational platforms allow for scaling analysis across space and time. Google Earth Engine (GEE) is an example of a platform that leverages cloud-computational services to achieve planetary-scale utility. GEE has been previously used for various research applications, including mapping population (Patel et al., 2015; Trianni et al., 2015), urban areas (Goldblatt et al., 2016) and forest cover (Hansen et al., 2013).

1.2. Research Objective

This paper develops a novel machine-learning methodology for supervised image classification of built-up areas that leverages high-resolution satellite data for analysis of large-scale regions using GEE`s cloud-based computational platform. Our methodology utilizes nighttime-light data as the source of training data for classification of built-up areas, yielding high precision without relying on expensive hand-labeled examples. It can be applied for any region on Earth.

Our methodology integrates two datasets: DSMP-OLS nighttime-light data and Landsat high-resolution daytime satellite imagery. We infer the spatial distribution of human activity and built-up land cover from nighttime-light data to collect examples for supervised image classification in Landsat's imagery. We assess the accuracy of the methodology using a dataset of 60,000 hand-labeled polygons characterizing built up (BU) and not built-up (NBU) pixels for each of the three study areas.

1.3. ***Study Area***

To illustrate our methodology and its applicability in heterogeneous and diverse geographical conditions, we map the built-up land cover in three countries that are characterized by distinct geographical conditions (e.g., land cover, topography, climate, soil, landform, and fauna): India, Mexico, and the US.

India. The share of India's population living in urban areas in 2015 was 33%, which is much lower than the corresponding values of 79% and 82% for Mexico and the US, respectively. However, India is urbanizing at a relatively rapid rate (Figure 1). For example, between 2010 and 2015, India's average annual rate of change of the urban population was 1.14%, compared to 0.36% and

0.21% in Mexico and in the US, respectively. By 2050, half of India's population is likely to be urban. In the last decade, the growth of India's urban population outpaced the growth of its rural population by 31.8% to 12.2% (H. S. Sudhira and K. V. Gururaja, n.d.)—primarily the result of natural urban population growth and secondarily because of rural-to-urban migration (Buhaug and Urdal, 2013)—a trend which is expected to continue (H. S. Sudhira and K. V. Gururaja, n.d.).

Mexico. Mexico has gone through three major phases of urbanization. In the first phase, 1900-1940, urban growth slowly incorporated 10% to 20% of the population. In the second phase, 1940-1980, rapid urban expansion, particularly in Mexico City, increased the share of urban population to 55%. In the current phase, since the 1980s, more dispersed moderate urban growth increased the urban share of the population to over 70% (Consejo Nacional de Población, 2012). The current phase of urbanization is characterized by informal urbanization on the city periphery, representing 65% of all new housing construction in Mexico City, and is even higher for small to mid-sized cities (Connolly, 2014). Informal settlements tend to be marginalized in terms of lower socio-economic development, access to services like water and electricity, and are more vulnerable to risks like water scarcity and flooding (Aguilar, 2008; Aguilar and Guerrero, 2013; Consejo Nacional de Población, 2012; Eakin et al., 2016).

US. In the context of the US, the initial urban growth occurred from 1790 to 1890 and the country has become increasingly urban since (Census Bureau, 2012). In 1910, the Census Bureau defined an urban area as one with a population above 2,500, and the 1920 census marked the first time that 50% of the US population lived in an urban area (US Census Bureau, 2016).

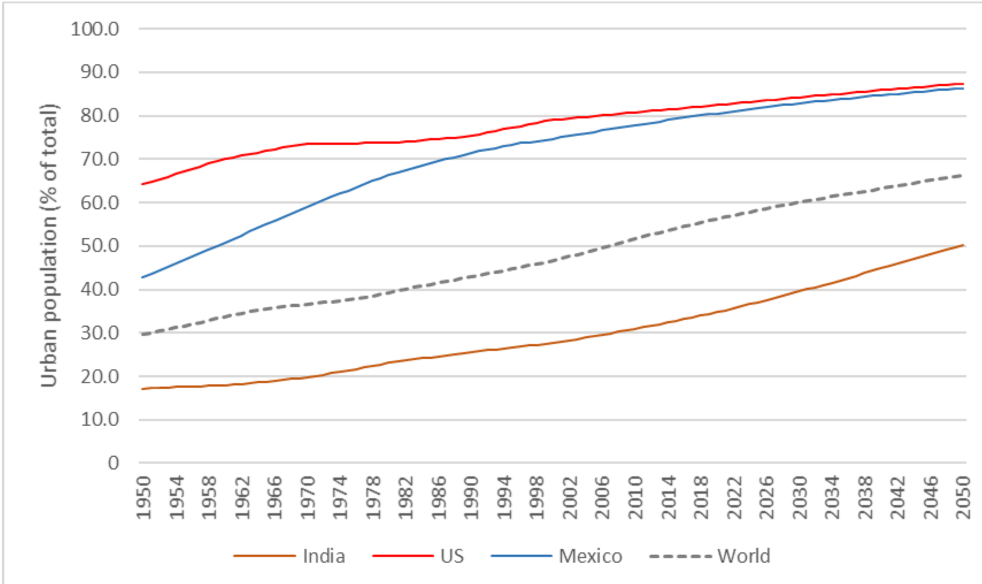

Figure 1: Annual changes in share of urban population in India, US and Mexico compared to world average (UN, 2014).

2. Conceptual Framework: Infused DMSP-OLS / Landsat Methodology

Our methodology relies on infusion of Earth-observation datasets from two domains: DMSP-OLS (which is used to extract examples of areas associated with human activity and built-up land-cover) and Landsat (which is used as the input for supervised image classification). We proceed in five steps (Figure 2): (1) Divide each country into a uniform hexagonal grid; (2) Pre-process Landsat 8 images; (3) Extract labeled examples from DMSP-OLS; (4) Perform supervised image classification; (5) Validate and test. We next describe these steps in detail.

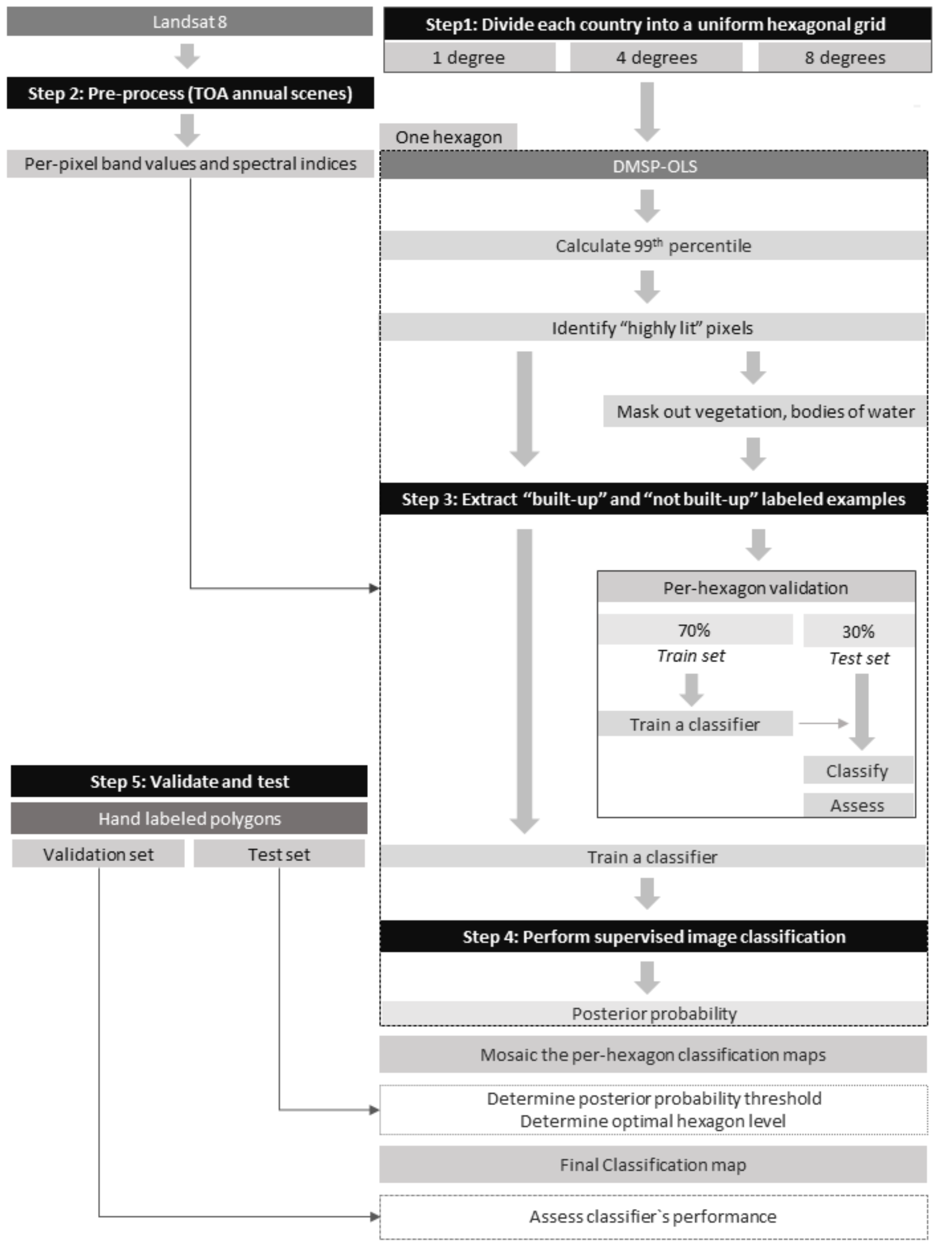

Figure 2: Schematic illustration of our infused DMSP-OLS / Landsat methodology.

2.1. Divide each country into an equal-area hexagonal grid ("mapping zones")

Mapping large-scale heterogeneous land cover requires partitioning the region of interest into a finite number of relatively homogenous sub-regions, or zones, that are characterized by similar landform, soil, vegetation, spectral reflectance, and image footprints (Homer et al., 2004). This practice is often referred to as 'zone mapping' (Homer and Gallant, 2001). The partition can be according to different criteria, such as land cover and land use, socio-political definition, size (Hunsaker et al., 1994; O'Neill et al., n.d.; Turner, 1989), or by means of an artificial grid system where each element in the grid is treated as an independent region of interest. In this study, we partition each country into an equal-area hexagonal grid (or a hexagonal tessellation). Hexagonal grids are advantageous because they are characterized by elements that do not have gaps or overlaps, the center-to-center distances between adjacent grid cells are approximately equal, the topology of the cells is symmetrical and invariant, the cells are equal area, and the cells can be recursively partitioned (Richards et al., 2000). Because the classification is subject to the size of each hexagon in the grid, we examine grids of different sizes of hexagons (different center-to-center distances): a distance of 1 decimal degree, 4 decimal degrees and 8 decimal degrees from center to center (see Figure 3 for illustration).

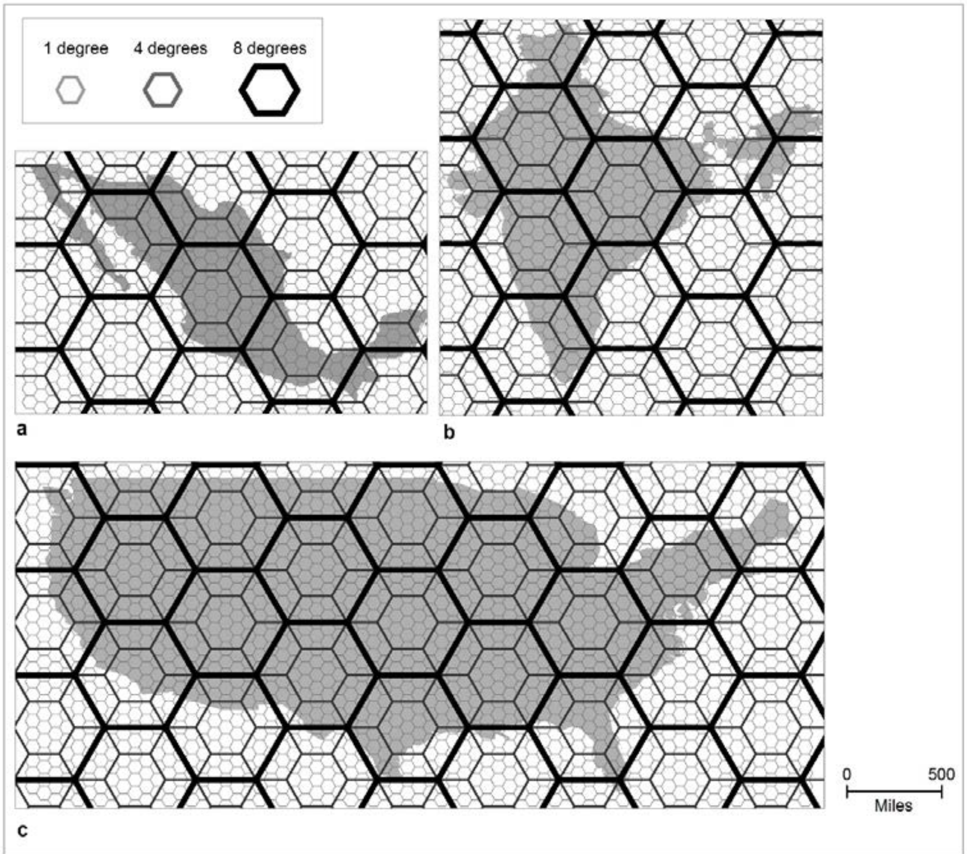

Figure 3: The three examined hexagon levels: 1°, 4°, and 8° from center to center, for (a) Mexico, (b) India, and (c) the US.

2.2. *Pre-process Landsat 8 images (classifier`s inputs)*

We use Landsat-8 imagery as classifier inputs (predictors). We apply a standard Top-of-Atmosphere (TOA) calibration on all USGS Landsat 8 Raw Scenes in one year (since DMSP-OLS is only available until 2013, we begin with mapping built-up areas in 2013). We assign a cloud score to each pixel and select the lowest possible range of cloud scores. Then we compute per-band percentile values from the accepted pixels and scale them to 8 bits. For each pixel we calculate additional spectral indices, which we use as additional predictors for the classifier: Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Urban Index (UI), Enhanced Vegetation Index (EVI), Normalized Difference Built-up Index (NDBI) (see description of these indices in Appendix 1).

2.3. Extract "built-up" and "not built-up" labeled examples from DMSP-OLS

Because "highly lit" pixels are associated with man-made structures that emit light, we assume that pixels with DN values that exceed a given threshold represent locations with built-up land cover and man-made structures. In this study we use DMSP-OLS (the "stable light" band of the 'F182013' satellite) to identify "highly lit" pixels. To account for regional variations, we determine this threshold for each hexagon independently by calculating the value of the 99th percentile of all pixels in the hexagon. A pixel is "built-up" if its DN value exceeds the threshold. Note that we only use hexagons that include at least one DMSP-OLS pixel with a value higher than 0. This definition allows us to capture, on the one hand, small settlements in isolated low-density regions (i.e., where the threshold is low), and on the other hand, only the core of cities in high-population-density regions (i.e., where the threshold is relatively high).

Due to the spatial resolution of DMSP-OLS and the blooming effect, areas identified as "highly lit" may potentially include non-built land cover. Thus, we also examine the effect of excluding these types of land cover from the lit pixels (according to Landsat's per-pixel NDVI and NDWI values). In each hexagon we randomly sample 100,000 pixels. We create a point at the center of each Landsat pixel and associate each point with the spectral values of the Landsat composite and derived spectral indices. Each example includes a label and the spectral values from Landsat. These per-hexagon training sets are used to build the local classification model.

2.4. *Perform supervised image classification*

As noted above, we train and classify each hexagon, with Random forest (20 trees) as the classifier. Random forests are tree-based classifiers that include k decision trees (k predictors). When classifying an example, its variables are run through each of the k tree predictors, and the k predictions are averaged to get a less noisy prediction (by voting on the most popular class). The learning process of the forest involves some level of randomness. Each tree is trained over an independently random sample of examples from the training set and each node's binary question in a tree is selected from a randomly sampled subset of the input variables. We use Random forest because previous studies find that the performance of Random forest is superior to other classifiers (Goldblatt et al., 2016), especially when dealing with large-scale and noisy datasets (Jean et al., 2016). We identify hexagons that remained unclassified as a result of no examples of built-up areas being sampled (e.g., hexagons that include only a few lit pixels, and that have a 99^{th} percentile value of DMSP-OLS pixels equal to 0). Because some of these hexagons do include isolated small settlements, which we want to capture, we additionally perform a classification using each country as one region of analysis (i.e., defining a single hexagon for a country) and use this classification to map the built-up land cover within these hexagons. Finally, we post process the classification maps by clipping the maps to the extent of the countries' borders. In addition, we remove misclassified built-up pixels in remote, unlit regions (i.e., where the DN value of the DMSP-OLS is 0).

2.5. *Validate and test*

We next assess the performance of the classifiers in each hexagon by dividing the sampled examples into a training set and a test set (30% and 70% of the examples, respectively). First, in each hexagon the classifiers are trained with the examples in the training set. We classify the examples in the test set and assess the classifier's performance. This validation procedure is designed to evaluate the quality of the sampled examples and to indicate how well our classifiers can predict their class. Then, we use the sampled examples for per-hexagon training and classification of built-up area. Finally, the classifications in all hexagons are mosaicked.

In each country we assess the accuracy of the classification using a large dataset of hand-labeled examples. We manually label these examples (polygons, 30m by 30m in size) as "built-up" or as "not built-up". We define polygons as built-up if the majority of their area (more than 50%) is paved or covered by human-made surfaces and used for residential, industrial, commercial, institutional, transportation, or other non-agricultural purposes. Similar definitions for urban areas are proposed by Goldblatt et al. (2016), Potere et al. (2009) and Schneider et al. (2010) who characterize a pixel as "urban" when the built environment spans the majority (50% or greater) of the sub-pixel space. For India, we use Goldblatt et al.'s (2016) ground-truth dataset. This dataset includes 20,030 examples (30m by 30m in size) labeled as "built-up" or as "not built-up" spanning the entire country (4682 polygons labeled as built-up and 16,348 labeled as not-built-up). For the US and Mexico, we construct a manually labeled ground-truth dataset of 20,000 examples (polygons, 30m by 30m in size) per country. Between 22%-27% of the datasets' polygons are labeled as built-up. We describe the procedure to create the stratified sample and the distribution of the examples in Appendix 2. We use half of the hand-labeled dataset to assess alternative parameters for the classifiers (the test set) and the other half to evaluate and report its performance (the validation set).

3. Results

3.1. Optimal Hexagon Scale and Parameters to the Classifiers

Our methodology produces high-quality, high-resolution maps of built-up areas for India, Mexico, and the US. First, we partition each country into a uniform hexagonal grid and consider each hexagon as an independent unit of analysis. To determine the optimal scale of the hexagons in the hexagonal grid division (1, 4, and 8 decimal degrees from center to center) and to assess various parameters to the classifiers, we evaluate the accuracy of the maps using half of the hand-labeled examples in each country (the test set). We use several performance estimators (we refer to the class "built up" as positive and to the class "not built-up" as negative): (1) True-Positive Rate (TPR) (the percentage of actual BU examples classified correctly as BU); (2) True-Negative Rate (TNR) (the percentage of actual NBU examples classified correctly as NBU); (3) Balanced Accuracy (the average of TPR and TNR); Precision (the percentage of actual BU examples out of all examples that were classified as BU) and F-Measure (the harmonic mean of the TPR and the precision):

1. TPR = TP / (TP + FN)
2. TNR = TN / (TN + FP)
3. Balanced Accuracy = (TPR + TNR) / 2
4. Precision = TP / (TP + FP)
5. F-Measure = 2 * ((Precision * TPR) / (Precision + TPR))

Where TP is the number of the actual BU examples predicted to be BU; TN is the NBU examples predicted as NBU; FN is the actual BU examples predicted as NBU and FP is the actual NBU examples predicted as BU.

The results show differences between countries in the optimal hexagonal scale. In India and Mexico, classification with the smallest level of hexagon (1°) results in the best performance, indicated by high balanced accuracy rates of 79% and 84%, respectively (Table 1), as well as by highest F-Measure scores (63% and 73%, respectively). Classification with larger hexagons results in a lower balanced accuracy: for example, with 8° hexagons it drops to about 75% and 77% in India and in Mexico, respectively. In the US, classification with the largest hexagons (8°) results in a marginally greater accuracy than with the smallest hexagon (1°), indicated by a balanced accuracy rate of 81.7% (compared to 81.5%). Classification with 8° hexagons also shows the highest F-measure value (67%). When no hexagons are used for classification and the classification is done with the entire country as one region of interest, both balanced accuracy and F-Measure drop. The classifiers predict for each new example (pixel) the probability it is a positive example ("built-up") (a posterior probability, ranging between 0 and 1). We find that the best performance in all three countries is achieved with a lower threshold on posterior probability (around 0.2). Table 2 presents, as an example, the performance of the classifiers as a factor of the posterior probability threshold, for classification with 1° hexagons. Both balanced accuracy and F-measure decrease as the posterior probability threshold increases. Although the TNR increases with higher posterior probability thresholds, the TPR decreases (i.e. the classification "misses" urban pixels). Thus, an optimal balanced accuracy is achieved with lower thresholds.

Table 1: Performance measures as a factor of the hexagon level for India, the US, and Mexico (classification with a posterior probability threshold of 0.1, 0.1 and 0.2, respectively).

		Overall accuracy	TPR	TNR	Balanced
India	no hexagon	80.5%	53.9%	88.2%	71.0%
	8°	78.1%	68.4%	80.9%	74.6%
	4°	80.8%	69.1%	84.2%	76.7%
	1°	80.3%	76.5%	81.4%	79.0%
US	no hexagon	82.4%	77.4%	83.9%	80.7%
	8°	81.5%	81.9%	81.4%	81.7%
	4°	74.2%	86.2%	70.6%	78.4%
	1°	78.0%	88.0%	75.0%	81.5%
Mexico	no hexagon	84.9%	46.1%	97.4%	71.8%
	8°	82.4%	66.2%	87.6%	76.9%
	4°	84.0%	71.0%	88.2%	79.6%
	1°	85.9%	79.9%	87.8%	83.8%

Table 2: The effect of the posterior probability threshold on the classifier's performance (classification using 1° hexagons).

	Threshold	0.1	0.2	0.4	0.6
India	Balanced	79.0%	75.6%	68.8%	63.6%
	F-Measure	63.2%	63.0%	53.3%	42.8%
US	Balanced	81.5%	80.2%	73.3%	67.1%
	F-Measure	65.0%	68.3%	61.0%	50.5%
Mexico	Balanced	81.4%	83.8%	78.6%	72.5%
	F-Measure	65.0%	73.2%	69.6%	61.1%

In the experiment described above, we sampled "built-up" examples from "highly lit" pixels (defined as all pixels with a DN value above the 99th percentile of the value across all pixels within a given hexagon). However, because these areas may also include other types of land cover, such as vegetation and bodies of water, in an additional experiment we mask out these types of land cover from the "highly-lit" pixels. We do this according to Landsat's per-pixel NDVI value (above 0.3 or 0.7) and NDWI values (a negative value). We find that excluding vegetation and bodies of water from the lit pixels does not affect much the performance of the classifiers (with 1° hexagons, performance only marginally improves by 0.2% and 0.6% in the US and Mexico, respectively). Thus, in the sequential experiments we do not remove these types of land cover from the "highly lit" pixels. Because the optimal hexagon level and posterior probability threshold varies between countries, we choose for each the following optimal parameters: for India 1° hexagons, posterior probability threshold 0.1; for Mexico 1° hexagons, posterior probability threshold 0.2; for the US 8° hexagons, posterior probability threshold 0.1.

3.2. Internal Per-Hexagon Accuracy Assessment

To evaluate the performance of the classifiers and the quality of our sampled training examples, we perform per-hexagon accuracy assessment (using only the sampled examples). This analysis is intended to assess the quality of the sampled examples and to provide an additional estimate on the performance of the classifiers in relation to spatial scale (it only relies on the sampled examples rather than on hand-labeled examples). In each hexagon, we randomly designate 70% of the examples for training and 30% for testing. We find a high balanced accuracy rate of 71.5% in India, 76.5% in the US. The balanced accuracy rate is lower in the case of Mexico (61%). These results indicate that the sampling procedure generates accurate examples that are beyond random. Although, by their nature, these examples are relatively "noisy", the classifiers can predict their class with reasonably high precision.

3.3. Accuracy Assessment with Validation Set

Based on the optimal hexagon level found for each country, we produce classification maps of built-up areas spanning the three countries. We use the second subset of our hand-labeled examples (the validation set) to assess the accuracy of the classification and find a high balanced accuracy rate of 79%, 80% and 84% in India, the US and Mexico, respectively. Similar to the accuracy measures found with the test set, the best performance is achieved with lower posterior probability thresholds, again, indicating a lower TPR as the posterior probability thresholds increase. Interestingly, while the TPR and TNR measures are relatively similar in India and the US (77% and 81%, and 79% and 81%, with a posterior probability of 0.1), in the case of Mexico the TPR is significantly higher than the TNR (92% and 71%, respectively, with a posterior probability of 0.1) (Table 3 presents the accuracy measures of the validation set and a confusion matrix of the classification performance).

Finally, we post-process the classification maps and clip them to the extent of each country. In addition, "low-lit" hexagons that include only a few lit pixels (i.e. where the 99th percentile of the lit pixels is 0) are mapped according to the classification we create for each country as one unit of analysis. We find that although the post-processing procedure only marginally improves the classification accuracy (see Appendix 3), a visual examination suggests that this procedure removes misclassified pixels, primarily over bodies of water and distant bare land. Thus, the final classification maps are post-processed.

Table 3: Description of accuracy measures and the confusion matrix using the validation set a (classification in India and Mexico is done with 1° hexagons, and in the US with 8° hexagons)

Posterior probability	Overall accuracy	TPR	TNR	Balanced	Precision	F-Measure					
					India						
0.1*	80.0%	77.0%	80.9%	79.0%	53.9%	63.4%			Confusion matrix		
0.2	83.9%	60.8%	90.6%	75.7%	65.2%	62.9%			Predicted		
0.3	84.0%	56.2%	92.0%	74.1%	67.0%	61.1%			BU	NBU	Sum
0.4	83.8%	44.7%	95.2%	69.9%	72.8%	55.4%	Actual	BU	1955	585	2540
0.5	83.5%	40.6%	95.9%	68.2%	74.3%	52.5%		NBU	1672	7101	8773
0.6	82.6%	32.2%	97.2%	64.7%	76.6%	45.3%		Sum	3627	7686	11313
					US						
0.1*	80.4%	78.7%	81.0%	79.8%	54.5%	64.4%			Confusion matrix		
0.2	84.9%	58.9%	92.4%	75.7%	69.2%	63.6%			Predicted		
0.3	84.8%	52.8%	94.0%	73.4%	71.9%	60.9%			BU	NBU	Sum
0.4	83.8%	38.2%	96.9%	67.6%	78.3%	51.4%	Actual	BU	1933	523	2456
0.5	83.3%	34.4%	97.4%	65.9%	79.4%	48.0%		NBU	1617	6873	8490
0.6	82.1%	26.3%	98.3%	62.3%	81.6%	39.8%		Sum	3550	7396	10946
					Mexico						
0.1	76.1%	91.7%	71.1%	81.4%	50.3%	65.0%			Confusion matrix		
0.2*	85.9%	79.9%	87.8%	83.8%	67.6%	73.2%			Predicted		
0.3	86.5%	75.4%	90.1%	82.7%	70.8%	73.0%			BU	NBU	Sum
0.4	86.8%	62.7%	94.4%	78.6%	78.2%	69.6%	Actual	BU	2123	534	2657
0.5	86.3%	57.4%	95.6%	76.5%	80.5%	67.0%		NBU	1018	7321	8339
0.6	85.2%	48.0%	97.1%	72.5%	83.9%	61.1%		Sum	3141	7855	10996

* denotes the highest balanced accuracy rate for which the confusion matrix is presented.

4. Discussion

We present a novel machine learning approach to map built-up areas at scale. Our methodology utilizes nighttime-light data (derived from DMSP-OLS) as a source for training examples of built-up and not built-up areas, which are then used for supervised image classification in Landsat 8 imagery. This is the first study, to our knowledge, to present a practical and simple form of transfer learning that can be applied to map built-up areas across space. Although many classification products map urban land, they are typically limited in their temporal and/or spatial resolution. This limits their use to track urbanization processes over time.

Mapping built-up areas at scale is challenging because of the scarcity of extensive ground-truth data for supervised classification and validation. Crowd-sourced datasets, such as OpenStreetMap (OSM) can also be used to map urban areas (Belgiu and Drăgu Painho, 2015). OSM is a valuable source for ground-truth data, primarily because of its vast extent and free availability. However, the completeness of OSM and its suitability for urban research is subject to the number and reliability of OSM contributors (Schlesinger, 2015). The use of OSM for supervised image classification remains challenging due to the risk of imbalanced distribution of class labels (including their spatial coverage), the presence of errors or missing class assignments ("class-noise"), and inaccurate polygon boundary delineations (Johnson and Iizuka, 2016). Our methodology overcomes the lack of such data by utilizing low-resolution DMSP-OLS data for classification of built-up areas in Landsat imagery. We collect examples of built-up and not-built-up areas by identifying "highly lit" areas within small homogenous regions (or mapping zones). These examples are used for image classification of built-up areas from Landsat imagery. By partitioning countries into smaller regions, we allow the parameters of the classification model to vary in what is determined as "built-up" pixels. We demonstrate that this flexibility is important and show that countries differ in this optimal hexagon scale. Although many studies address the effect of the classifiers` hyper-parameters on their performance, in this study we show that classifiers also have an optimal spatial scale, which can and should be discovered through techniques similar to those we propose here.

We assess the validity of our approach using two procedures for accuracy assessment: internal per-hexagon validation (assessing the classification of our sampled examples) and an external validation that uses 60,000 hand-labeled examples. The results demonstrate the robustness of our approach and its applicability in heterogeneous regions. We find that our classification performs well with a high balanced accuracy rate of around 80%. Yet, the degree to which the “localization” of our classifiers affects their performance varies between regions and depend on the heterogeneous nature of the mapped landscape. Appling this method at a global scale will require automatic methods for selecting, in any pixel, the scale of classification that maximizes accuracy. However, due to the lack of on-board calibration and unstable radiometric performance of the DMSP-OLS sensors, the absolute radiance of light cannot directly represent temporal changes in the intensity of the light, and thus, inter-sensor calibration is required to make our approach operational in time.

Our methodology overcomes the need for expensive hand-labeled data for supervised classification as well as many of the limitations associated with DMSP-OLS data. As illustrated in Figure 4, due to the blooming effect, lit areas are consistently larger than the built-up land cover they are associated with. Our classification captures the fine boundaries of built-up areas with high precision (Figure 5).

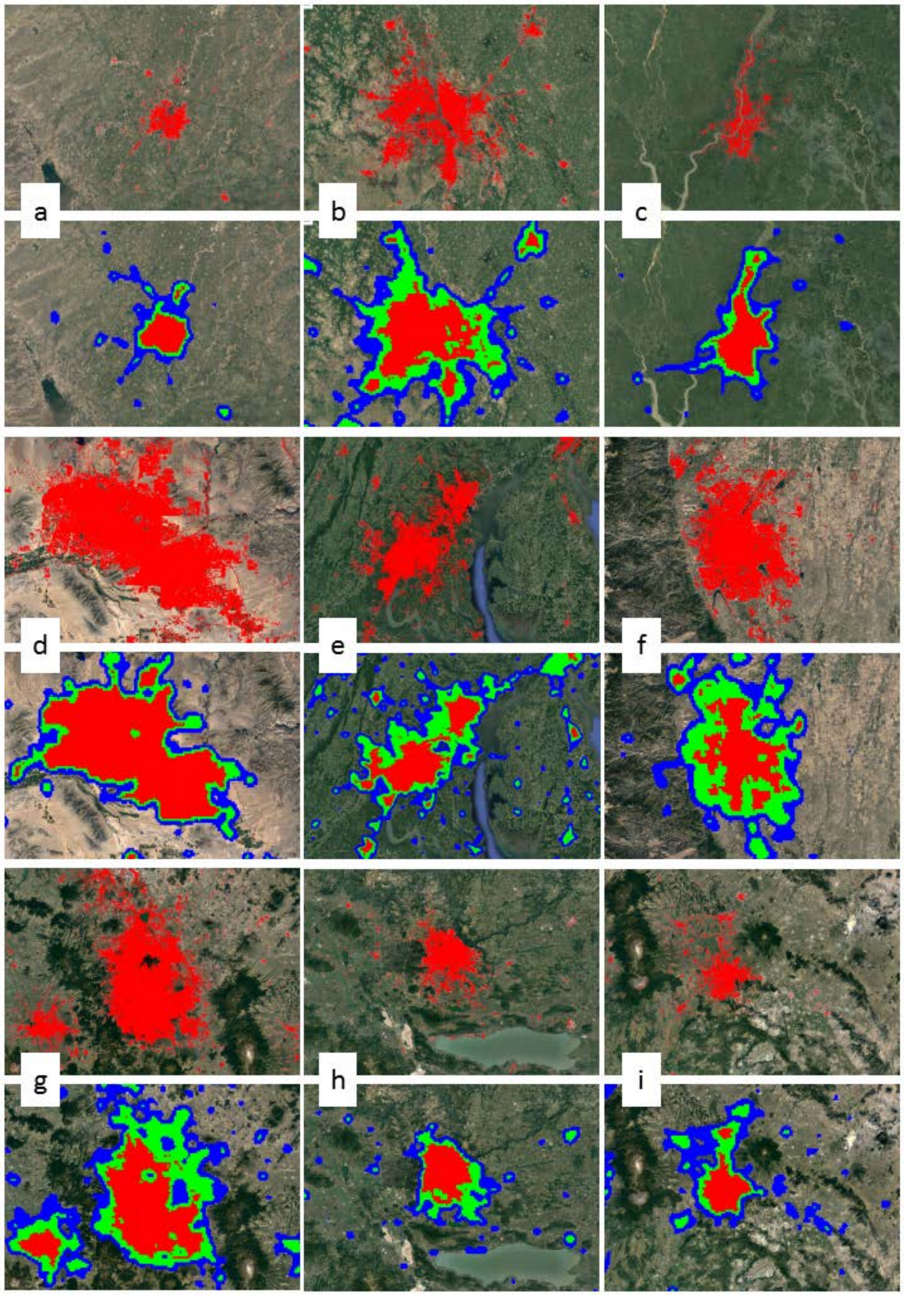

Figure 4: A comparison between our classification of built-up areas and lit pixels according to DMSP-OLS in (a) Ahmedabad, Gujarat, India; (b) New Delhi, Delhi, India; (c) Kolkata, West Bengal, India; (d) Phoenix, Arizona, US; (e) Washington DC, US; (f) Denver, Colorado, US; (g) Mexico City, Mexico; (h) Guadalajara, Mexico; (i) Puebla, Mexico. (the top figure in each city presents our classification; the bottom figure presents the DN values of DMSP-OLS stable lights band; 30-55 (red), 56-61 (green), 63 (blue))

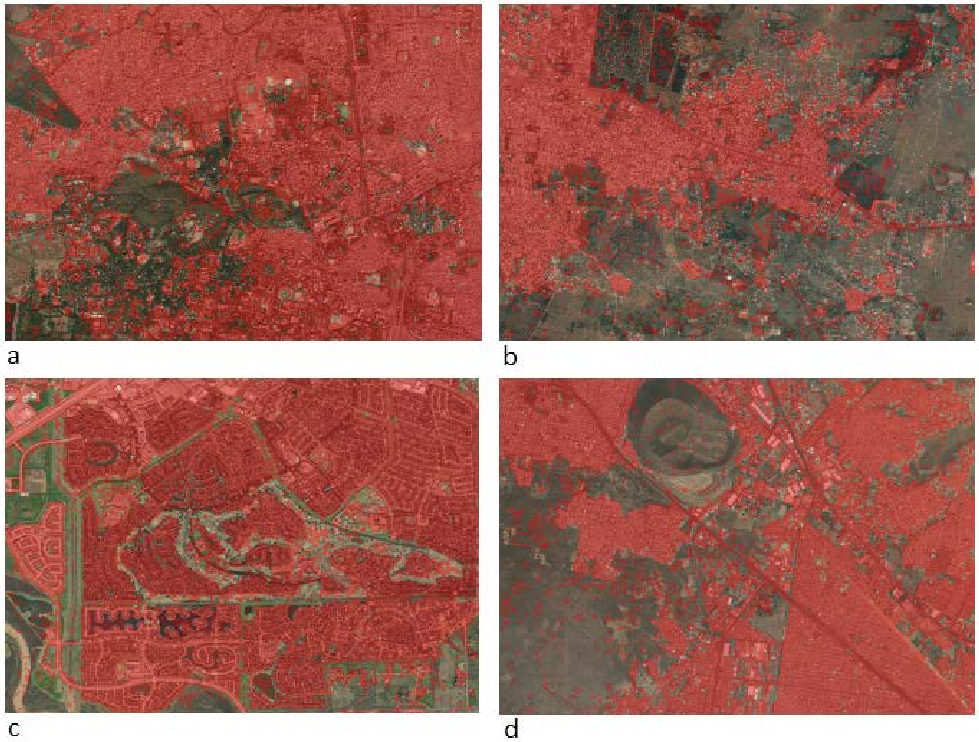

Figure 5: Classification of built-up areas (in red) in (a) Nagpur, Maharashtra, India; (b) Hyderabad, Telangana, India; (c) Houston, Texas, US; (d) Mexico city, Mexico

To evaluate our classification and to compare it with other products, we used the 60,000 labeled polygons to assess the accuracy of MCD12Q1 UMD MODIS classification scheme and DMSP-OLS (we define here a highly lit pixel similar to the definition we propose in this study, i.e., a value that is greater than the 99th percentile value of all pixels within a given hexagon). The accuracy of our classification exceeds both datasets; Table 4 shows that we achieve a higher balanced accuracy rate (between 6%-10%). As illustrated in Figure 6, the extent of the built-up land cover that we detect with our methodology is smaller than the extent of the land cover that is classified as urban and built-up by the MCD12Q1 classification.

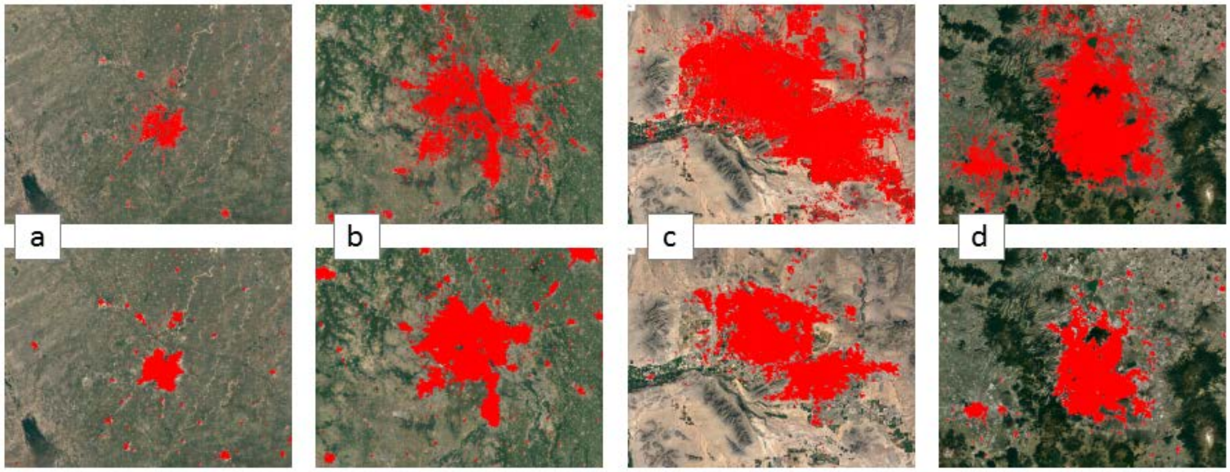

Figure 6: A comparison between areas classified as built-up using our methodology (top) and areas classified as built up and urban by MCD12Q1 UMD MODIS classification scheme (bottom), in (a) Ahmedabad, Gujarat, India; (b) New Delhi, Delhi, India; (c) Phoenix, Arizona, US; (d) Mexico City, Mexico

Our classification also exceeds other national high-resolution land-cover and land-use maps. To illustrate, a comparison between our examples and the US National Land Cover Database (NLCD) classification map showed a lower balanced accuracy rate in the NLCD product (72.3%) (we define a polygon as "built-up" if more than 50% of its area is built. Thus, we relate to the NLCD classes "Developed, Medium Intensity" and "Developed High Intensity" as "built-up"). Similar findings were found in the case of Mexico, where a comparison between our examples and the urban classification of the Instituto Nacional de Estadística y Geografía (INEGI) resulted in a lower balanced accuracy rate of around 79%.

Table 4: Accuracy assessment (balanced accuracy) of our infused methodology for classification, MODIS-MCD12Q1 (MODIS) and DMSP-OLS "highly lit" areas (defined as pixels with a DN value above the 99th percentile). Accuracy assessment with 60,000 labeled examples.

		Our BU classification	DMSP-OLS	MODIS UMD
India	TPR	74.9%	62.8%	61.7%
	TNR	81.8%	82.6%	84.3%
	Balanced	78.3%	72.7%	73.0%
US	TPR	87.5%	67.5%	64.4%
	TNR	74.8%	80.7%	87.3%
	Balanced	81.2%	74.1%	75.8%
Mexico	TPR	81.5%	77.6%	53.4%
	TNR	87.0%	71.2%	94.1%
	Balanced	84.3%	74.4%	73.7%

To summarize, we have developed a conceptual framework whereby utility of our transfer-learning methodology results in high-resolution, high quality depictions of built-up areas across three highly diverse countries. In today's era of big data, a globally consistent and data-driven method of defining and classifying urban areas has extensive applications. Economics, urban planning, climate modeling, water-resource management, hazard-response efforts, and urban-eco-system assessments all use geographic data on urban areas. With earth's rapidly urbanizing population, having information on urban extent that is spatially and temporally consistent and defined at high resolution is both relevant to a wide range of disciplines and essential for helping society better understand the drivers of urbanization.

5. References

Aguilar, A.G., 2008. Peri-urbanization, illegal settlements and environmental impact in Mexico City. Cities 25, 133–145. doi:10.1016/j.cities.2008.02.003

Aguilar, A.G., Guerrero, F.L., 2013. Poverty in peripheral informal settlements in Mexico City: The case of Magdalena Contreras, Federal District. Tijdschr. Voor Econ. En Soc. Geogr. 104, 359–378. doi:10.1111/tesg.12012

Bagan, H., Yamagata, Y., 2015. Analysis of urban growth and estimating population density using satellite images of nighttime lights and land-use and population data. GIScience Remote Sens. 52, 765–780. doi:10.1080/15481603.2015.1072400

Bagan, H., Yamagata, Y., 2014. Land-cover change analysis in 50 global cities by using a combination of Landsat data and analysis of grid cells. Environ. Res. Lett. 9, 064015. doi:10.1088/1748-9326/9/6/064015

Belgiu, M., Drăguţ, L., 2014. Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. ISPRS J. Photogramm. Remote Sens. 96, 67–75. doi:10.1016/j.isprsjprs.2014.07.002

Buhaug, H., Urdal, H., 2013. An urbanization bomb? Population growth and social disorder in cities. Glob. Environ. Change 23, 1–10. doi:10.1016/j.gloenvcha.2012.10.016

Census Bureau, U.S., 2012. Increasing Urbanization: Population Distribution by City Size, 1790 to 1890.

CIESIN, 2005. Gridded Population of the World, Version 3 (GPWv3) Data Collection.

Connolly, P., 2014. Latin Amercian Informal Urbanism: Contexts, Concepts, and Contributions with Specific Reference to Mexico, in: Becerra, A., Hernandez, F. (Eds.), (Re)viewing Latin American Cities. Cambridge Scholars Publishing, Cambridge.

Consejo Nacional de Población, 2012. Sistema Urbano Nacional 2012.

Eakin, H., Lerner, A.M., Manuel-Navarrete, D., Hernández Aguilar, B., Martínez-Canedo, A., Tellman, B., Charli-Joseph, L., Fernández Álvarez, R., Bojórquez-Tapia, L., 2016. Adapting to risk and perpetuating poverty: Household's strategies for managing flood risk and water scarcity in Mexico City. Environ. Sci. Policy. doi:10.1016/j.envsci.2016.06.006

Elvidge, C., Hsu, F.-C., Baugh, K., Ghosh, T., 2014. National Trends in Satellite-Observed Lighting: 1992-2012, in: Global Urban Monitoring and Assessment through Earth Observation, Remote Sensing Applications Series. CRC Press, pp. 97–120.

Estima, J., Painho, M., 2015. Investigating the Potential of OpenStreetMap for Land Use/Land Cover Production: A Case Study for Continental Portugal, in: Arsanjani, J.J., Zipf, A., Mooney, P., Helbich, M. (Eds.), OpenStreetMap in GIScience, Lecture Notes in Geoinformation and Cartography. Springer International Publishing, pp. 273–293.

Gaughan, A.E., Stevens, F.R., Linard, C., Jia, P., Tatem, A.J., 2013. High resolution population distribution maps for Southeast Asia in 2010 and 2015. PloS One 8, e55882. doi:10.1371/journal.pone.0055882

Georgescu, M., Chow, W.T.L., Wang, Z.H., Brazel, A., Trapido-Lurie, B., Roth, M., Benson-Lira, V., 2015. Prioritizing urban sustainability solutions: coordinated approaches must incorporate scale-dependent built environment induced effects. Environ. Res. Lett. 10, 061001. doi:10.1088/1748-9326/10/6/061001

Goldblatt, R., You, W., Hanson, G., Khandelwal, K.A., 2016. Detecting the Boundaries of Urban Areas in India: A Dataset for Pixel-Based Image Classification in Google Earth Engine. Remote Sens. 8. doi:10.3390/rs8080634

H. S. Sudhira, K. V. Gururaja, n.d. Population crunch in India: is it urban or still rural. Curr. Sci. 103, 37–40.

Hansen, M.C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 342, 850–853. doi:10.1126/science.1244693

Henderson, J.V., Storeygard, A., Weil, D.N., 2012. Measuring Economic Growth from Outer Space. Am. Econ. Rev. 102, 994–1028. doi:10.1257/aer.102.2.994

Henderson, M., Yeh, E.T., Gong, P., Elvidge, C., Baugh, K., 2003. Validation of urban boundaries derived from global night-time satellite imagery. Int. J. Remote Sens. 24, 595–609. doi:10.1080/01431160304982

Homer, C., Gallant, A., 2001. Partitioning the Conterminous United States into Mapping Zones for Landsat TM Land Cover Mapping. USGS.

Homer, C., Huang, C., Yang, L., Wylie, B.K., Coan, M., 2004. Development of a 2001 National Land-Cover Database for the United States. Photogramm. Eng. Remote Sens. 7, 829–840.

Hsu, F.-C., Baugh, K.E., Ghosh, T., Zhizhin, M., Elvidge, C.D., 2015. DMSP-OLS Radiance Calibrated Nighttime Lights Time Series with Intercalibration. Remote Sens. 7, 1855–1876. doi:10.3390/rs70201855

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213. doi:10.1016/S0034-4257(02)00096-2

Hunsaker, C.T., O'Neill, R.V., Jackson, B.L., Timmins, S.P., Levine, D.A., Norton, D.J., 1994. Sampling to characterize landscape pattern. Landsc. Ecol. 9, 207–226. doi:10.1007/BF00134748

Imhoff, M., 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 Sens. Environ. 61, 361–370. doi:10.1016/S0034-4257(97)00046-1

Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. Science 353, 790–794. doi:10.1126/science.aaf7894

Johnson, B.A., Iizuka, K., 2016. Integrating OpenStreetMap crowdsourced data and Landsat time-series imagery for rapid land use/land cover (LULC) mapping: Case study of the Laguna de Bay area of the Philippines. Appl. Geogr. 67, 140–149. doi:10.1016/j.apgeog.2015.12.006

Kawamura, M., Jayamanna, S., Tsujiko, Y., 1996. Relation between social and environmental conditions in Colombo Sri Lanka and the urban index estimated by satellite remote sensing data. Int. Arch. Photogramm. Remote Sens. 31, 321–326.

Keola, S., Andersson, M., Hall, O., 2015. Monitoring Economic Development from Space: Using Nighttime Light and Land Cover Data to Measure Economic Growth. World Dev. 66, 322–334. doi:10.1016/j.worlddev.2014.08.017

Levin, N., Duke, Y., 2012. High spatial resolution night-time light images for demographic and socio-economic studies. Remote Sens. Environ. 119, 1–10. doi:10.1016/j.rse.2011.12.005

Liu, Y., Delahunty, T., Zhao, N., Cao, G., 2016. These lit areas are undeveloped: Delimiting China's urban extents from thresholded nighttime light imagery. Int. J. Appl. Earth Obs. Geoinformation 50, 39–50. doi:10.1016/j.jag.2016.02.011

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432.

Miyazaki, H., Iwao, K., Shibasaki, R., 2011. Development of a New Ground Truth Database for Global Urban Area Mapping from a Gazetteer. Remote Sens. 3, 1177–1187. doi:10.3390/rs3061177

O'Neill, R. V., Hunsaker, C.T., Timmins, S.P., Jackson, B.L., Jones, K.B., Riitters, K.H., Wickham, J.D., n.d. Scale problems in reporting landscape pattern at the regional scale. Landsc. Ecol. 11, 169–180. doi:10.1007/BF02447515

Pan, S.J., Yang, Q., 2010. A Survey on Transfer Learning. IEEE Trans. Knowl. Data Eng. 22, 1345–1359. doi:10.1109/TKDE.2009.191

Patel, N.N., Angiuli, E., Gamba, P., Gaughan, A., Lisini, G., Stevens, F.R., Tatem, A.J., Trianni, G., 2015. Multitemporal settlement and population mapping from Landsat using Google Earth Engine. Int. J. Appl. Earth Obs. Geoinformation 35, Part B, 199–208. doi:10.1016/j.jag.2014.09.005

Pesaresi, M., EHRLICH Daniele, FERRI Stefano, FLORCZYK Aneta, CARNEIRO FREIRE Sergio Manuel, HALKIA Stamatia, JULEA Andreea Maria, KEMPER Thomas, SOILLE Pierre, SYRRIS Vasileios, 2016. Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014. Publications Office of the European Union, Ispra (VA), Italy.

Pettorelli, N., Vik, J.O., Mysterud, A., Gaillard, J.-M., Tucker, C.J., Stenseth, N.C., 2005. Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends Ecol. Evol. 20, 503–510. doi:10.1016/j.tree.2005.05.011

Potere, D., Schneider, A., Angel, S., Civco, D.L., 2009. Mapping urban areas on a global scale: which of the eight maps now available is more accurate? Int. J. Remote Sens. 30, 6531–6558. doi:10.1080/01431160903121134

Richards, T., Gallego, J., Achard, F., 2000. Sampling for forest cover change assessment at the pan-tropical scale. Int. J. Remote Sens. 21, 1473–1490. doi:10.1080/014311600210272

schlesinger, J., 2015. Using Crowd-Sourced Data to Quantify the Complex Urban Fabric - OpenStreetMap and the Urban–Rural Index, in: Arsanjani, J., Zipf, A., Mooney, P., Helbich, M. (Eds.), OpenStreetMap in GIScience. Springer International Publishing, Switzerland, pp. 295–315.

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 Sens. Environ. 114, 1733–1746. doi:10.1016/j.rse.2010.03.003

Seto, K.C., Fragkias, M., Güneralp, B., Reilly, M.K., 2011. A Meta-Analysis of Global Urban Land Expansion. PLOS ONE 6, e23777. doi:10.1371/journal.pone.0023777

Seto, K.C., Güneralp, B., Hutyra, L.R., 2012. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. Proc. Natl. Acad. Sci. 109, 16083–16088. doi:10.1073/pnas.1211658109

Small, C., Elvidge, C.D., 2013. Night on Earth: Mapping decadal changes of anthropogenic night light in Asia. Int. J. Appl. Earth Obs. Geoinformation 22, 40–52. doi:10.1016/j.jag.2012.02.009

Small, C., Pozzi, F., Elvidge, C.D., 2005. Spatial analysis of global urban extent from DMSP-OLS night lights. Remote Sens. Environ. 96, 277–291. doi:10.1016/j.rse.2005.02.002

Stevens, F.R., Gaughan, A.E., Linard, C., Tatem, A.J., 2015. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. Plos One 10, e0107042. doi:10.1371/journal.pone.0107042

Su, Y., Chen, X., Wang, Chongyang, Zhang, H., Liao, J., Ye, Y., Wang, Changjian, 2015. A new method for extracting built-up urban areas using DMSP-OLS nighttime stable lights: a case study in the Pearl River Delta, southern China. GIScience Remote Sens. 52, 218–238. doi:10.1080/15481603.2015.1007778

Sutton, P.C., 2003. A scale-adjusted measure of "Urban sprawl" using nighttime satellite imagery. Remote Sens. Environ., Urban Remote Sensing 86, 353–369. doi:10.1016/S0034-4257(03)00078-6

Tatem, A.J., Noor, A.M., Hagen, C. von, Gregorio, A.D., Hay, S.I., 2007. High Resolution Population Maps for Low Income Nations: Combining Land Cover and Census in East Africa. PLOS ONE 2, e1298. doi:10.1371/journal.pone.0001298

Taubenböck, H., Esch, T., Felbier, A., Wiesner, M., Roth, A., Dech, S., 2012. Monitoring urbanization in mega cities from space. Remote Sens. Environ. 117, 162–176. doi:10.1016/j.rse.2011.09.015

Trianni, G., Lisini, G., Angiuli, E., Moreno, E.A., Dondi, P., Gaggia, A., Gamba, P., 2015. Scaling up to National/Regional Urban Extent Mapping Using Landsat Data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 3710–3719. doi:10.1109/JSTARS.2015.2398032

Turner, M.G., 1989. Landscape Ecology: The Effect of Pattern on Process. Annu. Rev. Ecol. Syst. 20, 171–197.

UN, 2014. World Urbanization Prospects: The 2014 Revision (No. ST/ESA/SER.A/352).

US Census Bureau, 2016. Urban and Rural Areas [WWW Document]. URL https://www.census.gov/history/www/programs/geography/urban_and_rural_areas.html

Wei, Y., Liu, H., Song, W., Yu, B., Xiu, C., 2014. Normalization of time series DMSP-OLS nighttime light images for urban growth analysis with Pseudo Invariant Features. Landsc. Urban Plan. 128, 1–13. doi:10.1016/j.landurbplan.2014.04.015

Zha, Y., Gao, J., Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. Int. J. Remote Sens. 24, 583–594. doi:10.1080/01431160304987

Zhang, Q., Seto, K.C., 2013. Can Night-Time Light Data Identify Typologies of Urbanization? A Global Assessment of Successes and Failures. Remote Sens. 5, 3476–3494. doi:10.3390/rs5073476

Zhou, Y., Smith, S.J., Elvidge, C.D., Zhao, K., Thomson, A., Imhoff, M., 2014. A cluster-based method to map urban area from DMSP/OLS nightlights. Remote Sens. Environ. 147, 173–185. doi:10.1016/j.rse.2014.03.004

Zhou, Y., Smith, S.J., Zhao, K., Imhoff, M., Thomson, A., Bond-Lamberty, B., Asrar, G.R., Zhang, X., He, C., Elvidge, C.D., 2015. A global map of urban extent from nightlights. Environ. Res. Lett. 10, 054011. doi:10.1088/1748-9326/10/5/054011

Appendix 1: Indices used in Machine Learning Classifier

I. NDVI (Normalized Difference Vegetation Index)

NDVI expresses the relation between red visible light (which is typically absorbed by a plant's chlorophyll) and near-infrared wavelength (which is scattered by the leaf's mesophyll structure) (Pettorelli et al., 2005). It is computed as:

(NIR-RED) / (NIR+RED) (1)

where NIR is the near infra-red wavelength and RED is the red wavelength. The values of NDVI range between (-1) and (+1). An average NDVI value in 2014 was calculated for each pixel (with Landsat 7 32-Day NDVI Composite).

II. NDBI (Normalized Difference Built-up Index)

NDBI expresses the relation between the medium infra-red and the near infra-red wavelengths (Zha et al., 2003). It is computed as:

(MIR-NIR) / (MIR+NIR) (2)

where MIR is the medium infra-red and NIR is the near infra-red wavelength. The index assumes a higher reflectance of built-up areas in the medium infra-red wavelength range than in the near infra-red.

III. NDWI (Normalized Difference Water Index)

NDWI expressed the relation between the green (G) and the NIR (near infra-red), with a scaling of -1 to +1(McFeeters, 1996). It is computed as:

(G-NIR) / (G+NIR) (3)

The positive values are typically open water areas with reflect green light but not NIR wavelengths, while negative values are non-water features, like soil and vegetation, which reflect higher NIR values than green wavelengths.

IV. UI (Urban Index)

UI is the difference between the short infra-red (SWIR) and the near-infrared wavelengths (Kawamura et al., 1996). It is computed as:

(SWIR-NIR) / (SWIR+NIR) (4)

Like NDBI, UI assumed high brightness in SWIR in urban areas as opposed to the NIR.

VI. EVI (Enhanced Vegetation Index)

EVI is an improved vegetation index with higher sensitivity in high biomass regions where NDVI tends to saturate, reduces atmospheric influences, and removes the canopy background signal(Huete et al., 2002). It is computed as:

(2.5 * ((NIR-R) / (NIR + 6R – 7.5B +1)) (5)

This is a similar formula to the NDVI, which takes advantage of high reflectance of vegetation in the NIR band as opposed to the R band. The blue band (B) is used (with a coefficient of 7.5) to correct for aerosol influences in the red band (R). There is a integer of 1 added to the denominator to adjustment for nonlinear NIR and R radiant transfer through canopies. 2.5 is applied as a gain to the index.

Table 6. Bands in landsat and indices used

	Spectral band	Wavelength (micrometers)	Resolution (meters)
	Landsat 8		
B1	Band 1 – Ultra blue	0.43 - 0.45	30
B2	Band 2 - Blue	0.45 - 0.51	30
B3	Band 3 - Green	0.53 - 0.59	30
B4	Band 4 – Red	0.64 - 0.67	30
B5	Band 5 - Near Infrared (NIR)	0.85 - 0.88	30
B6	Band 6 - SWIR 1	1.57 - 1.65	30
B7	Band 7 - SWIR 2	2.11 - 2.29	30
B8	Band 8 - Panchromatic	0.50 - 0.68	15
B10	Band 10 - Thermal Infrared (TIRS) 1	10.60 - 11.19	100 (resampled to 30)
B11	Band 11 - Thermal Infrared (TIRS) 2	11.50 - 12.51	100 (resampled to 30)
NDVI	(B5-B4)/(B5+B4)		30
NDWI	(B3-B5)/(B3+B5)		30
NDBI	(B6-B5)/(B6+B5)		30
EVI	2.5*((B5/B4)/(B5+6*B4-7.5*B2+1)		30
UI	(B7-B5)/(B7+B5)		30

Appendix 2: Sampling Scheme for Testing and Validation

We use a random stratified sampling procedure according to the nighttime light intensity. In each country, we identify the NTL (night time lights) pixels whose value is 63 (the highest possible value). Then we calculate a 5 pixels radius circle-shaped boolean kernel (a buffer of approximately 5 Km). We found this method to result in an approximate distribution of 75% urban points, our class of interest, and 25% non-urban points. We randomly sample ~10,000 points in this buffer zone of high NTL and its periphery. This process was repeated twice, to generate two 10,000 point datasets for each country- only used in testing to determine optimal algorithm parameters, and the second used to validation the final urban class map to assess its accuracy.

A 30 by 30 meter square buffer (the size of a Landsat pixel) is drawn around each random sample point. These polygons are overlaid with high resolution imagery in Google Earth. The interior of each polygon is compared to the imagery it overlays by an analyst. Based on the imagery, the polygons are manually labeled as built-up or as not-built-up. A polygon is built-up when 50% or more of the contents of polygon are man-made. Polygons labeled not built-up may still contain man-made structures, such as roads or buildings, but they make up less than 50%.

Table 7: the distribution of the built-up and not-built-up points for each country.

	BU points	NBU points	Total
India	4682 (22.26%)	16,348 (77.74%)	21,030
US	4386 (21.6%)	15,898 (78.4%)	20,284
Mexico	5477 (27.4%)	14,523 (72.6%)	20,000

Appendix 3: Accuracy Assessment of Post Processed Country Maps

All samples (~20,000 per country) described in appendix 2 were used to assess accuracy when adding in pre-processing and after post-processing. Pre-processing involved replacing hexagons with low NTL in each country with the urban classification map produced without hexagons (e.g. using the entire country boundary to extract training data and build a classifier). This is because hexagons with low NTL do not have enough training data to make a good classification. Post-processing includes the pre-processing steps in addition to removing pixels with 0 NTL values. Results are shown in table 8. India has marginal accuracy improvements, Mexico shows no effect, and the US improves accuracy around 1%. Table 3 shows the results for various accuracy metrics.

Table 8: Accuracy assessment pre- and post-processing (removing pixels with 0 NTL values)

	India		Mexico		US	
	Pre-Process	Post- Process	Pre-Process	Post- Process	Pre-Process	Post- Process
Accuracy =	0.807372748	0.810091743	0.862046781	0.862046781	0.819810714	0.833104772
Precision =	0.54927557	0.554166061	0.735567091	0.735567091	0.575346505	0.600739372
Recall =	0.751963865	0.748428908	0.779723195	0.779723195	0.789239482	0.786448881
TPR =	0.751963865	0.748428908	0.779723195	0.779723195	0.789239482	0.786448881
TNR =	0.823244824	0.827733453	0.893364628	0.893364628	0.828795625	0.846781344
Balanced =	0.787604345	0.788081181	0.836543912	0.836543912	0.809017553	0.816615113
F-Measure =	0.634833789	0.636811764	0.757001781	0.757001781	0.665529592	0.681163217