

Equitable development through deep learning: The case of sub-national population density estimation

Patrick Doupe
Arnhold Institute for Global
Health
New York, NY, USA
patrick.doupe@mssm.edu

James Faghmous
Arnhold Institute for Global
Health
New York, NY, USA
james.faghmous@mssm.edu

Emilie Bruzelius
Arnhold Institute for Global
Health
New York, NY, USA
emilie.bruzelius@mssm.edu

Samuel G. Ruchman
Office of the UN
Secretary-General's
Special Envoy for Health in
Agenda
2030 and for Malaria
New York, NY, USA
sruchman@healthenvoy.org

ABSTRACT

High-resolution population density maps are a critical component for global development efforts, including service delivery, resource allocation, and disaster response. Traditional population density efforts are predominantly survey driven, which are laborious, prohibitively expensive, infrequently updated, and inaccurate – especially in remote areas. Furthermore, these maps are developed on a regional basis where the methods used vary region to region, hence introducing notable spatio-temporal heterogeneity and bias.

The advent of global-scale satellite imagery provides us with an unprecedented opportunity to create inexpensive, accurate, homogeneous, and rapidly updated population maps. To fulfill this vision, we must overcome both infrastructure and methodological obstacles. We propose a convolutional neural network approach that addresses some of the methodological challenges, while employing a publicly available, albeit low resolution, remote sensed product. The method converts satellite images into population density estimates. To explore both the accuracy and generalizability of our approach, we train our neural network on Tanzanian imagery and test the model on Kenyan data. We show that our method is able to generalize to unseen data and we improve upon the current state of the art by 177 percent.

1. INTRODUCTION

The 2015 adoption of the Sustainable Development Goals has created an acute need to collect accurate subnational statistics to monitor progress towards these targets [14].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM DEV '16, November 17-22, 2016, Nairobi, Kenya

© 2016 ACM. ISBN 978-1-4503-4649-8/16/11...\$15.00

DOI: <http://dx.doi.org/10.1145/3001913.3001921>

Furthermore, subnational data are a critical component of equitable development since national averages routinely “smooth out” the most marginalized communities. In fact, one of the most significant limitations to achieving the Millennium Development Goals was that, while national averages may have showed overall gains towards certain goals, these gains veered significant disparities, where the poorest people did not necessarily benefit from the overall progress [38].

One of the most basic, yet essential, of these statistics is the accurate and timely assessment of where people live. This information is “one of the primary sources of data needed for formulating, implementing and monitoring the effectiveness of policies and programmes aimed at inclusive socioeconomic development and environmental sustainability” [40, p.2]. These census data determine resource allocation, such as where to invest in hospitals, schools and infrastructure, and may be used to define legislative districts and other important functional areas of government. Basic population level data also serve as an essential benchmark for measuring progress toward the attainment of the Sustainable Development Goals, and other national and international objectives (see Figure 1).

Traditional population density estimates are derived from census surveys. However the utility of censuses is hampered by several widely recognized quality issues. First, they are “among the most complex and massive peacetime exercises a nation undertakes” [42, p.5]. Many countries still lack the capacity, both in terms of financial and human resources, to collect data regularly [7]. As a result, population maps in many low income countries are outdated or of poor quality [27]. For instance, in several low- and middle-income counties a full census is conducted less than once per decade [15]. Infrequent data collection limits the utility of censuses.

A second concern is problematic data quality stemming from biased reporting and incomplete coverage. Low data quality is most concerning from a development point of view since many of the most vulnerable people are also those most likely to be undercounted [44]. For example, until 2014 Myanmar had not completed a census in nearly two

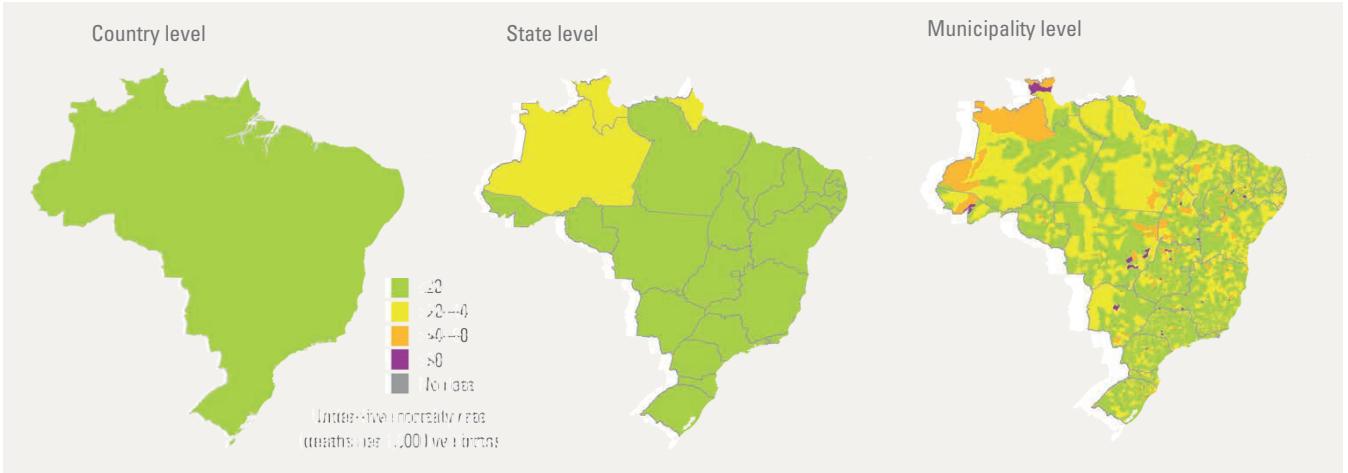


Figure 1: Disaggregated data reveals subnational disparities, enabling targeted policy interventions and more efficient, equitable resource allocation. E.g., Brazil’s national under-five mortality rate was 16 deaths per 1000 live births in 2013. However, more than 30 municipalities suffered an under-five mortality rate at 80 deaths or more per 1000 live births. Many of these municipalities are located in states with low mortality rates. [1]

decades. The population of a northern region not controlled by the central government, subsequently determined to be nearly half a million people, had never been counted at all [41]. Current estimates suggest that 300–350 million people worldwide may not be included in official country population figures [6]. Poorly estimated population figures result in poor decision-making. For example, in Africa, correcting poor estimates translates to a 50 percent increase in the population of the lowest income quantile without access to an improved water source [6].

Last, national level data obscures important subnational heterogeneity. Understanding subnational heterogeneity in preferences and constraints is required for subnational planning and policy delivery. Effective planning and policy delivery requires spatially disaggregated data, which are often not be available at the desired local level [38]. For example, accurate estimates of a population-at-risk for disease requires accurate, geographically explicit population data [32]. This data does not always conform to the specific administrative boundaries of census data. Amongst others, the resulting error has consequences for public health delivery. Hay *et al.* note that depending on which map we use to estimate malaria risk in Kenya, the number of people at risk may vary by up to 1 million [20]. Globally, estimates may vary by up to 10 million [33].

Figure 1 shows an example of national aggregate data obstructing subnational disparities. The left panel shows national aggregate data for under-five mortality rate in Brazil. Looking at national aggregates alone, one would think Brazil is doing well in this department. But as we disaggregate the data into finer spatial resolutions (middle and right panels) we notice significant health outcome disparities as well as significant data sparsity.

Two major survey programs provide countries with technical and financial assistance to assess a variety of population and health data: the USAID Demographic and Health Survey (DHS) program and the UNICEF Multiple Indicator Cluster Survey (MICS) program. Standard DHSs are typi-

cally conducted every five years, though some countries have not historically conducted these regularly [22]. Since 2009, UNICEF has provided countries with support to implement MICS every three years if countries so decide [39]. However, once surveys are conducted, it can take several years before final data are released, limiting capacity for any rapid response to recently updated data.

Survey costs run into the millions for each survey. Recent DHS total costs have ranged from USD 1.1 million to 9.7 million per survey: 40 percent of DHSs cost between USD 3.0 million to 5.9 million and 20 percent fall in the upper bracket, costing USD 6 million to 9.7 million total due to large sample sizes (and complex biomarkers) [45]. In Europe, census costs have risen considerably, topping a median cost of USD 5.57 per capita in 2010 for traditional censuses. The largest share of these costs were incurred in the field [3]. For surveys conducted in low-income countries, the crucial task of creating sampling frames is particularly costly. For example, the MICS method requires that all households in a sample cluster (e.g., a village, census enumeration area, etc.) be listed in order to generate a final list of households for surveying. To reduce bias, the manual listing exercise is executed separately from surveying, doubling costs by requiring that clusters be visited twice [37]. Accurate pre-existing household maps could reduce survey costs and further reduce bias in sampling within selected clusters. Moreover, identifying previously un- or under-counted populations could improve the accuracy of nationally representative surveys.

The standard non-survey method of mapping populations is to distribute the population estimate for a specified area over the geographic space according to a set of rules [5, 13]. However, the rules used are rarely optimized to minimize mapping error. Even if optimized to minimize error [30], the rules are often defined based on modifying satellite images in combination with a range of other data sources. Both of these steps introduces further possibilities for mapping error, limiting the accuracy of maps.

Figure 2 demonstrates some of these practical limitations. The left panel shows a high-resolution image of Naivasha, Kenya. The middle panel shows population density estimates of the region from CIESIN (Columbia University). Unfortunately, the spatial scale of these estimates are too coarse to guide any meaningful policy or decision-making. Effective policy requires high-resolution, frequently updated, and accurate estimates such as the high-resolution estimates produced by Facebook in the right panel.

In this paper we propose a new method that uses a Convolutional Neural Network (CNN) to overcome many of these limitations. The CNN generates rules for distributing population estimates [25]. The rules are constructed to minimize the error between estimate and actual population densities. Error is also reduced by removing intermediate data creation steps: raw satellite images are used with only basic modifications. This method improves upon the state of the art by 177 percent.

CNNs typically perform well when identifying objects within images because they retain local spatial information. This capability is useful for mapping population levels, because human activities typically result in landscapes that are spatially distinct, and therefore recognizable, from ‘natural’ settings. For example, the straight lines of transportation networks and buildings are associated with higher population density. Rural regions like mountains have a less rigid spatial structure and are generally associated with lower population density. Figures 3 and 4 show the features extracted from satellite sensors to train the CNN. The features are from images for an urban and a rural area in Kenya. We see that the urban area is more fragmented, whereas the rural area has smoother gradation of colours. The difference in features provides us with confidence that a CNN will distinguish dense from more sparsely populated areas.

CNNs are mainly used for object classification, but have been also used for regression problems. A common classification tasks is identifying whether an image is of a cat or a dog. This is modelled as choosing between a 1 (for cat) or 0 (for dog). CNNs perform best on such tasks. Where CNNs have been used in regression tasks, where the aim is to estimate a number like the population of a region. Example tasks include estimating human poses from photographs [36] and bounding boxes for text [23].

Our choice of data has benefits too. Since the only inputs are satellite images, up-to-date information can be accessed each time new satellite images become available. Since high quality satellite images have global coverage, this method can be used to estimate population anywhere on earth. The high frequency, low resolution data and more accurate model enables more accurate prevalence mapping for a variety of important policy, planning and monitoring tasks.

2. RELATED WORK

In this section we present the applied context of our work. For additional reviews, see [10, 20, 5].

The most basic method for constructing population maps is to simply assume constant population density across a region. Using this method, the average population count for Kenya would be 78 persons per 1km^2 land area. Of course, we know there is a higher population density in some regions (e.g., Nairobi) and less in others (e.g., Marsabit County). Therefore if we wanted to preserve some heterogeneity at the subnational level, we could collect population counts for dis-

tricts and apply the same method. This common approach, known as *areal weighting*, was used to develop the first version of the Gridded Population of the World (GPW) map [34]. Although areal weighting is sufficient for many tasks, it leaves sharp differences in population estimates across regional boundaries, which rarely occur in human settlements.

Interpolation methods [35] estimate unknown data points for specific locations using known data values in other locations. These methods are used to smooth population estimates across administrative boundaries. Additional information can also be used to improve estimates. For instance, although we only have population information down to a given sub-national unit, if we know that one cell within the unit contains a city and another cell does not, this additional information can be used to reallocate the estimates. For example, the Global Rural Urban Mapping Project (GRUMP) uses additional information from night time light images to detect urban and rural areas [4]. LandScan [26] is a highly granular population dataset that uses road and land cover information to reallocate population estimates. The UN Environmental Program has also used similar methods to construct high resolution spatial maps for Africa, Asia and Latin America [11]. Similarly, AfriPop, AsiaPop and LatinPop [46] use urban and rural as well as land class information to further refine spatial estimates. Population weights are determined by the average population in the class.

There are two groups that use machine learning to set weights. Stevens et al. [30] use a random forest model. They use a wide variety of input data layers. This model improves on the past literature, with lower mean absolute errors (MAE) and root mean square errors (RMSE) than GPW, GRUMP and AsiaPop. The second group is the Facebook Connectivity Lab [16]. This group also uses a CNN, but uses this to identify buildings. Population is then distributed equally among the buildings. We instead directly estimate population.

The difference between these maps and ours are the inputs and the weighting scheme. The inputs are generated by third parties for other purposes. For instance, WorldPop uses data on the extent of cities, slope of mountains and waterbodies. As with all datasets, these arrive with error. It is uncertain how these errors combine through to the final estimate. Further, there may be inputs available that are either not included or not yet developed. This missing data contributes to error in the final estimate. The weighting scheme in the literature implies uniform population distribution within a land class. Because a discrete set of inputs are used, there is a discrete set of population classes. This assumption creates error since the population distribution within a land class may not be uniform [31].

We present work that builds on these previous approaches while overcoming certain limitations. First, we construct a model for weighting various population cells based on satellite images alone, avoiding the middle man of constructing variables. If our model is good enough, we will be able to capture the information contained in the input layers created by other authors. This provides an additional benefit: the possibility of capturing information available from satellite images yet not existing as input layers. Since we don’t have land classes, we don’t assume uniform distribution within land classes. Also, the method has the capacity to directly estimate population without a census count, all that is required is a satellite image and a trained model.

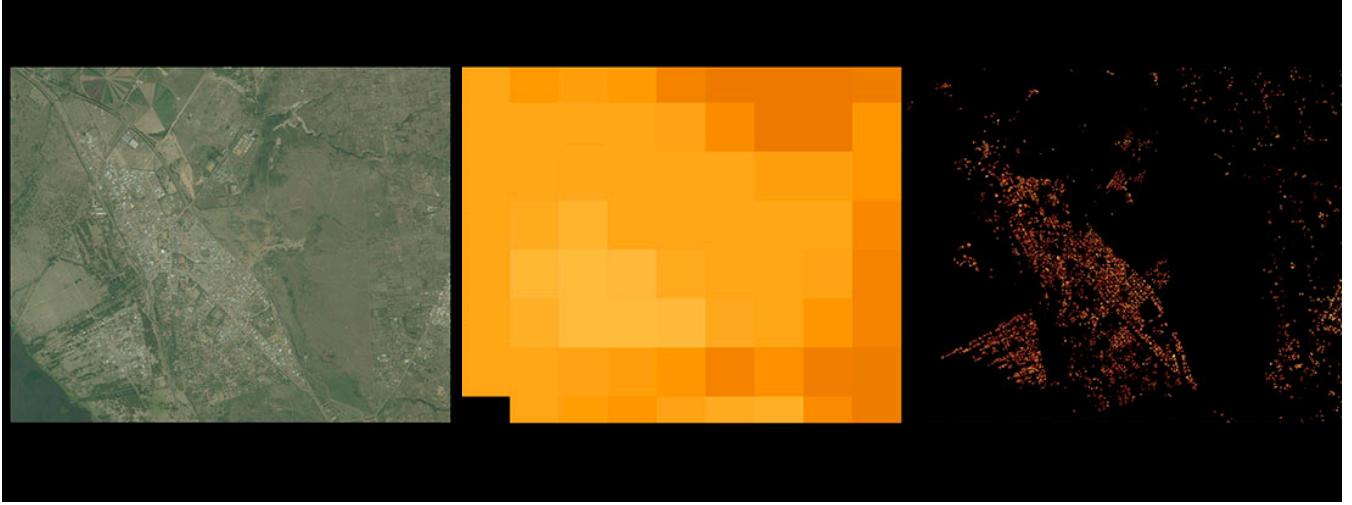


Figure 2: Highlighting the limitations of traditional population density estimation techniques. Most population density estimates are too coarse to have any operational value (middle panel). Effective policy requires high-resolution, frequently updated, and accurate estimates (right panel). Left: a high-resolution image of Naivasha, Kenya. Middle: A gridded population density estimate of the region from CIESIN (Columbia University). Right: A high-resolution population density map from Facebook’s proprietary population density estimation algorithm. Source: Facebook and Digital Globe.

Figure 3: Features extracted from satellite images of a populated area in Nairobi, Kenya. Subtitles refer to image bands.

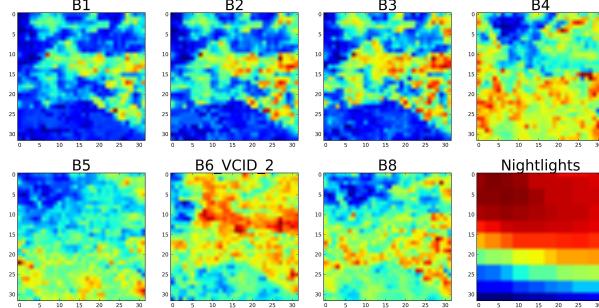
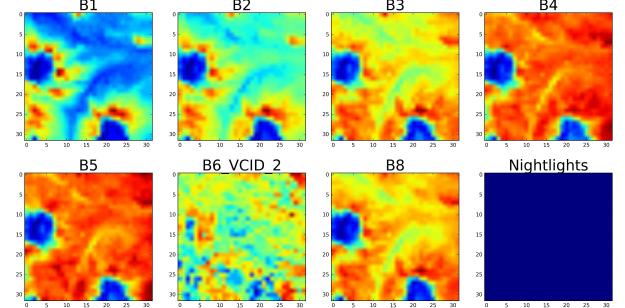


Figure 4: Features extracted from satellite images of a rural area near Marsabit, Kenya. Subtitles refer to image bands.



3. DATA

We use two types of data: satellite data and population data. The satellite data were drawn from two different sources, LANDSAT 7 mission images [2] and Defense Meteorological Satellite Program Operational Line Scanner (DMSP/OLS) night time lights [18] data. For both sources of data we collected images from a three year window surrounding the census year. For instance, Tanzanian census data was obtained for 2002, so the corresponding satellite data coverage was from 2001 to 2003. Google Earth Engine was used to download images.¹

The satellite data.

We begin with orthorectified raw LANDSAT 7 images over the three year window for both nations. Cloudy images (with a cloud score greater than 1) were removed from

the collection. Seven spectral bands were selected.² In addition to standard red, green and blue lights, thermal infrared and other wavelengths were used to allow us to investigate spatial patterns not obvious to the human eye. A Top-Of-Atmosphere composite of was taken for each band [8]. We downloaded images with pixel sizes of approximately 250m².

For the DMSP/OLS data, we used the ‘stable lights’ band. This band reflects a cleaned version of the raw data, with ephemeral events like fires removed from the dataset. This leaves us with indicators of human activity. Although the data is modified, it was included because the quality is globally uniform and has a long time series. Several authors have suggested that nighttime light may have less utility for population density estimates of African nations [21]. Nevertheless we included them within these analyses as they provided some utility in for identifying major cities. We

¹<http://code.earthengine.google.com>

²Specifically, B1, B2, B3, B4, B5, B6 VCID2 and B8. The B7 band was omitted due to data limitations.

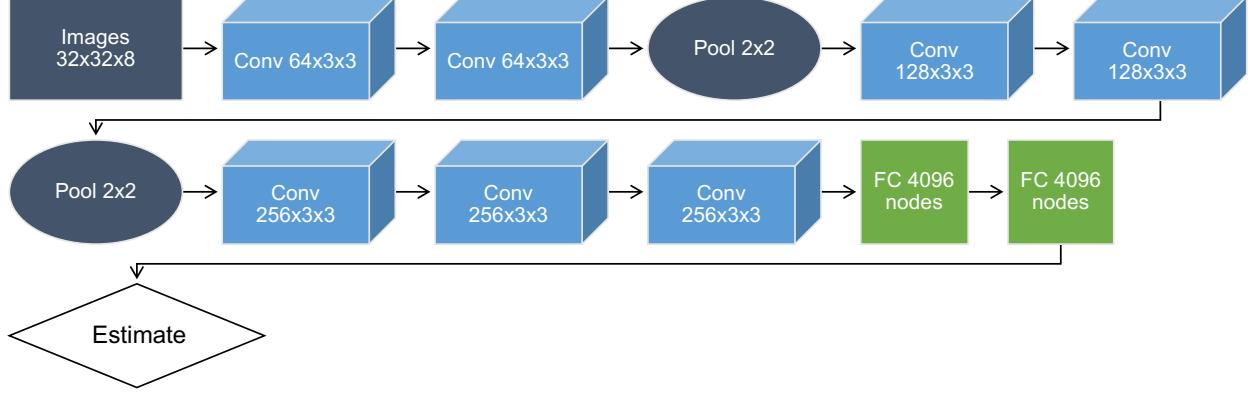


Figure 5: We use three main types of layers to build our architecture. Convolutional Layer (Conv), Pooling Layer (Pool), and Fully-Connected (FC) Layer. All Conv layers have zero padding and use rectified linear unit (ReLU) layer activation. Pool layers use max pooling

downloaded images with pixel sizes of approximately 1km².

Figures 3 and 4 show the utility of including night lights in our study. There is no value for any pixels in the Marsabit nightlight band (Figure 4), whereas for Nairobi the image is activated (Figure 3). Further, we can see the difference in information across the bands. This visual inspection justifies their inclusion in the analysis. All bands were downloaded as geoTIFF files, permitting the conversion of pixel positions to georeferenced coordinates.

The population data.

We obtained 2002 census data for Tanzania and 2009 census data for Kenya. The spatial resolution is Enumeration Area for Tanzania and the Sublocation level for Kenya. From here on, we will also refer to either the Enumeration Area or the Sublocation as *regions*. These were the highest level spatial resolution we could obtain for each country. There are 18421 Tanzanian Enumeration Areas. For Kenya, we have 7150 Sublocations. We merged the Kenyan data to a shapefile of jurisdictional boundaries. The shapefile is from 1999 and due to changing Sublocation boundaries, we have 6622 Sublocations in our merged dataset.

The seven year period between Tanzanian and Kenyan census data collection introduces some noise into our system. Our model maps a relationship between what satellites can see and population estimates. Where this relationship changes over time, then a model trained on data in one period will not predict as well in later periods. An example of this would be large scale urban-rural migration of people, with little changes in the building stock. We do not suspect this will affect the example in this paper, we would warn against the use of this method to predict the amount of people in large scale migrations, rural-urban or otherwise.

Combining the datasets.

There are three components to the dataset creation. First we merge the satellite and population datasets. This step assigns population to each pixel in the satellite images. Second we create *observations*, small images, for model training. Last, we take a weighted random sample of the data to ensure we have a good amount of observations with positive

populations. All work was done using Python libraries.³

We begin by merging the population and satellite data. For each pixel in the satellite data, we assigned the population density of the region the pixel's coordinates were located in. For cells that covered regional boundaries, the cell's density value was determined by the region of the cell's centroid.

To generate observations for model use, we sliced the large image into 32 × 32 × 8 pixel observations. The window size was chosen to be 32 × 32 because of the large pixel size. With approximately 250m² pixels, a 32 × 32 yields a 8km² window. A VGG-net size image of 224 × 224 pixels would yield a window of 54km². Only 222 windows would be sufficient to cover Kenya. Even with flipping and rotating overlapping images, the dataset size would be small. The cost of this design is that we train our model from scratch, rather than using pretrained weights.

Observations were taken by a sliding window (or convolution) procedure. The window started in the first row and column and sliced an observation. Then the window shifted right by 8 pixels and took another 32 × 32 × 8 slice. When the last column was reached, the window shifted down by 8 pixels, moved to the first column and started again. This resulted in a geographically uniform distribution of observations.

To generate observations for model use, we sliced the large image into 32 × 32 × 8 pixel observations. Observations were taken by a sliding window (or convolution) procedure. The window started in the first row and column and sliced an observation. Then the window shifted right by 8 pixels and took another 32 × 32 × 8 slice. When the last column was reached, the window shifted down by 8 pixels, moved to the first column and started again. This resulted in a geographically uniform distribution of observations.

The resulting population density distribution followed a power law. The power law arises because large areas of land have small population densities and only a small total area has large population densities. If we used the sliced observations or a random sample, the model would be trained to

³See <http://github.com/ArnholdInstitute/ACM-DEV> for code and libraries.

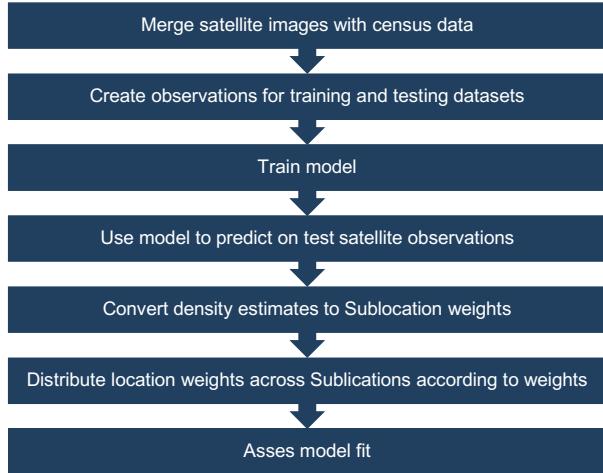


Figure 6: The path from satellite images to final predictions.

recognise areas with low densities and fail to pick up areas with high densities. To obtain a workable distribution of population density, we selected a random sample of observations weighted by population density.

The last step was to normalise the data. We log normalised the output data (see Figure 4). For the input data, we subtracted each pixel for each band by the band’s mean for that pixel and scaled by the band’s standard deviation for that pixel. In the final training dataset, 145649 observations were included. For Kenyan test data, we followed a similar procedure but did not take a weighted random sample.

4. METHOD

We present a methodology to estimate population from satellite images (see figure 6). Although we focus on Tanzanian 2002 and Kenyan 2009 censuses, this method is broadly applicable. All data and software are available for free in open-source format at: *redacted for blind review*

Once we have constructed our data, our method follows a four step process. First we train the model using Tanzanian data. Second, we estimate population density for arbitrary satellite images for Kenya. Next, we convert these estimates into weights. Last, we use these weights to distribute known population at one regional level amongst lower regional levels. This procedure of estimating weights and distributing population data is common in the literature. Since we use LANDSAT satellite images, we call this method *LANDSAT landstats (LL)*.

We also present results that use the first two of the above steps. That is, we take the outputs from the CNN as population density estimates. This allows us to estimate population without census data. To distinguish this method from the distributed method, we call this method *LL-raw* and the distributed method *LL-distributed*.

Model training.

We base our model architecture (Figure 5) on the well known VGG-net [29]. Our deviate from VGG-net to accommodate our input image size. As we explained in Section 3, our input size is restricted by the relatively large pixels in our

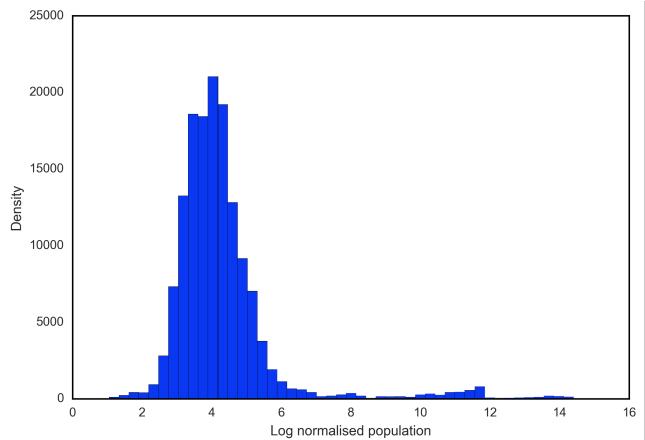


Figure 7: Histogram of log normalised population used in training observations.

dataset. We cross validated against layer size. The model was trained with mini batch gradient descent and backpropagation using the Adam optimizer [24]. The model was trained using Keras [9].

We chose the VGG-net model since it has high performance in image classification and a simple architecture. We begin with two convolutional layers and two fully connected layers. With each of the convolutional layers were rectified linear unit activation layers (ReLU). We also had a max-pooling layer after the two convolutional layers. We then tested the adding of additional layers against performance on a cross validated set (20 percent of training set). Performance gains after eight convolutional layers was minimal.

During training we augmented the input data. Since population density is the same for a city whether it faces north, south, east or west, we could randomly rotate our images 90, 180 and 270 degrees without imposing bias. By augmenting our data, we increased the effective size of our training set by four.

Estimation.

After training, we used the model to predict population densities for Kenyan observations. These observations were normalised using the parameters from training data. Observations were passed through the model once to obtain an estimate of population density. We assigned this density value to a center pixel. Since the observations are 32 by 32 pixels, there is no single center pixel. Rather, we took a pixel at row 17, column 17. We end up with a regular grid of points across Kenya.

We converted these density estimates to Sublocation values by taking the mean value of all points within a Sublocation. This means that for an observation with a city on the boundary of two regions the model will assign the whole density to the region where the observations’ center pixel coordinates lie. An alternative approach would be to apportion density to both regions, however we choose the former as our estimates are fine grained enough for this to be a reasonable approximation.

Estimating population.

To construct these weights for LL-distributed, we first take the above estimate for Sublocation s in location ℓ , $\bar{d}_{s,\ell}$. Given this estimate, and a Sublocation's area, we can estimate population $n_{s,\ell}$. Next, we obtained the estimated population share of a Sublocation within a location $w_{s,\ell}$. The last step is to distribute a location's population p_ℓ given these weights.

$$\hat{p}_{s,\ell} = w_{s,\ell} p_\ell$$

where

$$w_{s,\ell} = \frac{n_{s,\ell}}{\sum_{t \in S_\ell} n_{t,\ell}}$$

$$n_{s,\ell} = \bar{d}_{s,\ell} \text{Area}_{s,\ell}$$

LL-raw takes the population estimates $n_{s,\ell}$ as the population estimate. This method is by definition less likely to be accurate, but we calculate it for two main reasons. First, understanding how the model generates raw estimates can help us improve the distributed method. Second, raw estimates can be used in the absence of census data.

4.1 Error metrics

Following [30] we use three metrics to measure performance of predicted population at the sublocation level. The first is the root mean squared error RMSE, defined in the usual way. The benefit of the RMSE is that it squares deviations from accuracy. Therefore, higher deviations from accuracy result in a more than higher error score. Because this is not as interpretable, the percent mean squared error %RMSE is also presented. The %RMSE is the RMSE divided by the average actual Sublocation population. This provides a measure of how far off a prediction is from the average population. Last is the mean absolute error MAE. The MAE is that average deviation of a sublocation's estimated population from its actual population.

5. CHALLENGES

There are three main challenges to this method. As ours is an exploratory analysis, we erred on the side of simplicity in resolving these challenges. Scaling up this analysis in future work will require careful thought to overcome these challenges.

The first challenge, as mentioned in Section 3, is that we have not fully established the external validity of the method. We assumed that spatial patterns of human settlement are similar between Tanzania in 2002 and Kenya in 2009. If these patterns are similar enough, what about training a model on 1992 Tanzanian data? Or using data from Nigeria or India in 2002? How many CNNs will be required to get a global estimate? The answers are not obvious. Scaling up this method will require testing along both time and space dimensions to validate this claim [17].

Second, the scale of the data is large. We limit our initial analysis to two nations. Further, we did not choose a 'dense' grid of all possible observations nor did we choose the highest possible pixel resolution. Increasing the scale along these three dimensions is desired: more nations means more coverage and greater generalizability; a denser grid of observations means more observations for training; a higher

resolution means buildings and roads can be directly observed. Increasing along these three lines will increase the performance of the model, but there will be costs of database creation and design.

Last, optimizing the architecture of the CNN will require additional work. Many of the rules of thumb for CNN design are based upon RGB images used for a classification problem. We have 8 band images used for a regression problem. We chose to use a small model based on the VGG net as a starting point.

An alternative option was to use a pretrained models. Despite the promise others have shown that pretrained models have with satellite images [47]. We opted against this for two related reasons. First, the image size of VGG net is 224x224 pixels, seven times larger than our images. With such image sizes and pixel resolutions around 250m across, we would have observations around 56km^2 . Such large observations may smooth out heterogeneity in population density. An aim of this research is to generate high resolution maps of population. With our still coarse resolution, we opted for smaller image sizes. Second, we use eight bands of information. Image classification CNNs would use three bands. Using pretrained networks means we would not use valuable information [43].

6. RESULTS

The LL-distributed improves upon the current state of the art's RMSE by 177 percent. Table 1 shows the RMSE is around 2102 with an average error of around 1038 people per Sublocation. The MAE is around 1038 people per sublocation. At an average population per sublocation of 4083, we have an average error rate of 25% of the population. A scatter plot of the estimates for each image against the actual population in the area the image shows the error does not seem to change by population size (see Figure 8).

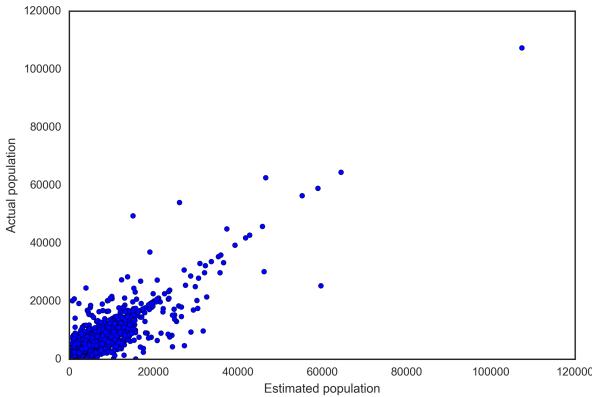
Table 1: The relative performance of LANDSAT-landstats methods against other methods. Values for other methods taken from [30].

Method	RMSE	% RMSE	MAE
LL-distributed	2102.36	51.48	1037.82
Forest et al	3956.74	91.36	1685.82
AsiaPop	5208.79	120.28	2184.64
LL-raw	5938.24	145.43	3963.17
GRUMP	6294.61	145.35	2383.64
GWP	6327.80	146.11	2304.70

Figures 9 and 10 show the estimated population, actual population and estimation error across Sublocations in the Coast Province of south east Kenya. As Table 1 also shows, the model does well at matching the population. We see higher populations along the south west and south east borders. The model also estimates well the lower population Sublocations in the center and north of the Province. From figure 9, it seems that the model overestimates low population areas and underestimates high population areas.

The model is also capable of generating raw estimates for population from satellite observations. The performance of the model falls; however, the RMSE is better than well

Figure 8: Estimated Kenyan Sublocation population against actual population.



known methods like GRUMP and GWP. Comparing the MAE shows the limits of the current model. The MAE is around 166 percent higher for the LL-raw estimates. In general, the LL method model is noisy, but picks up relative differences amongst nearby regions.

7. DISCUSSION AND CONCLUSION

The CNN-based population mapping method presented here provides an enhanced approach for estimating population distributions at sub regional levels. This information is necessary for informed planning and policy efforts. Our main method LL-distributed improves upon the current state of the art. The method yields a 177 percent increase in accuracy. For our additional method, LL-raw that estimates population without distributing census data, performance is comparable to existing approaches.

Our method overcomes several well described data quality issues. The input data – satellite data – is available globally at uniform quality (conditional on cloud coverage) and spatially refined levels. Because the satellite data is rapidly updated, and the method does not require hand generated layers, updated population data sets can be generated in a much more timely and effective manner compared to many other methods. In the development of this research, we have seen improvements prediction quality with higher resolution data. As high resolution satellite imagery is becoming increasingly available, the method we have proposed represents an accurate and low-cost complement or alternative to conventional population mapping approaches.

Despite these strengths of the method, there are several limitations that should also be highlighted. First, as with the majority of CNN approaches, the method is a ‘black box’ meaning that it has somewhat limited interpretability. In comparison, existing population mapping methods generally provide some intuition as to where the model suggests people live [30]. This can help build a narrative of human settlement. Whilst there has been some work to understand what factors of inputs drive predictions in a CNN, interpreting CNNs is still an open issue.

Another important limitation is that CNNs are likely to have blind spots. For example, if a region has a unique or rare settlement pattern, then the model is unlikely to recognise these qualities. Further, vacant buildings are indistinguishable from occupied buildings from space. The same

idea extends to situations where the population of buildings exhibit seasonality; think of resorts in winter or office buildings at night. In terms of development, these blind spots are an important concern because cities in rapid transition may not be well estimated. Other methods, such as those that use mobile phone user data, may be more useful in settings where rare settlement patterns are known or suspected [19, 28, 12, 48].

Although our algorithm can provide useful predictions without any labeled data, its use within a development setting presents further opportunities for improved performance and utility. Conditional on having a trained model and the necessary satellite images, the method has the potential to provide population change measures for any region over any given time. It is also possible to modify the output variable to other examples where data are available. For instance, the method might additionally be used to map population poverty levels, certain health indicators or other important population characteristics [47]. One direction for future work would be to link survey data to examine the more complex relationships and characteristics of populations that are needed to address outstanding policy and planning questions. These benefits suggest this method can help us better understand of societies change, and better plan for the future.

8. REFERENCES

- [1] Adapted from UNICEF, Committing to child survival: A promise renewed, Progress Report 2015 (UNICEF: New York, 2015), 74 (figure 44). Source data from Brazilian Ministry of Health, Department of Informatics, DATASUS.
- [2] Landstat7. <http://science.nasa.gov/missions/landsat-7/>. Accessed: 2016-06-22.
- [3] Measuring population and housing: practices of the unicef countries in the 2010 round of censuses. Technical Report 3, United Nations Statistics Division, 2014.
- [4] D. Balk, F. Pozzi, G. Yetman, U. Deichmann, and A. Nelson. The Distribution of People and the Dimension of Place: Methodologies to Improve the Global Estimation of Urban Extents. In *Urban Remote Sensing Conference*, 2005.
- [5] D. L. Balk, U. Deichmann, G. Yetman, F. Pozzi, S. I. Hay, and A. Nelson. Determining Global Population Distribution: Methods, Applications and Data. In A. G. a. D. J. R. Simon I. Hay, editor, *Advances in Parasitology*, volume 62 of *Global Mapping of Infectious Diseases: Methods, Examples and Emerging Applications*, pages 119–156. Academic Press, 2006.
- [6] R. Carr-Hill. Missing millions and measuring development progress. *World Development*, 46:30–44, 2013.
- [7] R. O. Censuses et al. Advocacy and resource mobilization. 2005.
- [8] G. Chander, C. Huang, L. Yang, C. Homer, and C. Larson. Developing consistent landsat data sets for large area applications: The mrlc 2001 protocol. *IEEE Geoscience and Remote Sensing Letters*, 6(4):777–781, 2009.
- [9] F. Chollet. Keras. <https://github.com/fchollet/keras>,

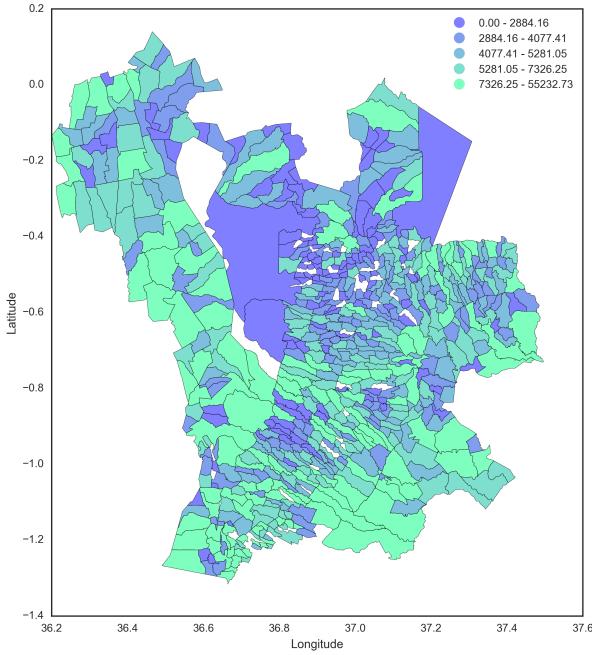


Figure 9: Estimated sublocation populations

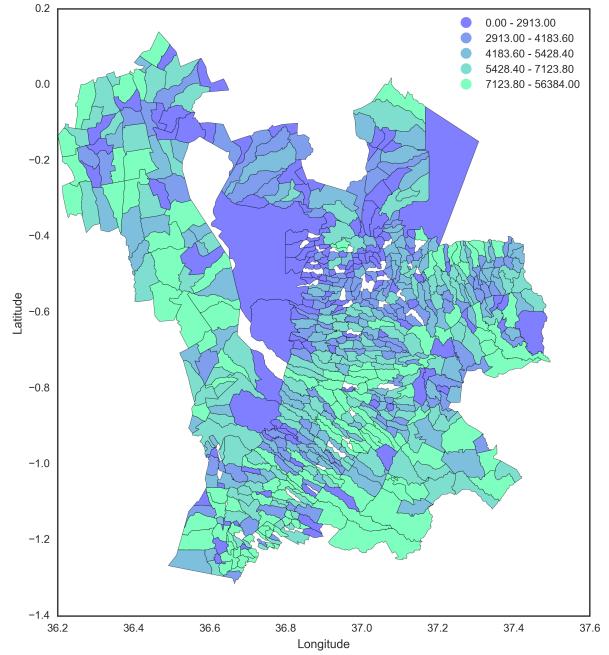


Figure 10: Actual sublocation populations

2015.

- [10] U. Deichmann. A Review of Spatial Population Database Design and Modeling (96-3). *eScholarship*, Mar. 1996.
- [11] U. Deichmann and L. Eklundh. *Global digital datasets for land degradation studies : a GIS approach*. United Nations Environment Programme, Nairobi, 1991.
- [12] P. Deville, C. Linard, S. Martin, M. Gilbert, F. R. Stevens, A. E. Gaughan, V. D. Blondel, and A. J. Tatem. Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*, 111(45):15888–15893, Nov. 2014.
- [13] P. J. Diggle and E. Giorgi. Model-based geostatistics for prevalence mapping in low-resource settings. *Journal of the American Statistical Association*, (just-accepted):1–42, 2015.
- [14] U. N. S. Division. Better data better lives: 47th session (2016). <http://unstats.un.org/unsd/statcom/47th-session/documents/>. Accessed: 2016-06-28.
- [15] U. N. S. Division. Census dates for all countries. <http://unstats.un.org/unsd/demographic/sources/census/censusdates.htm>. Accessed: 2016-06-28.
- [16] Facebook Connectivity Lab. Connecting the world with better maps, 2016.
- [17] J. H. Faghmous and V. Kumar. Spatio-temporal Data Mining for Climate Data: Advances, Challenges, and Opportunities. In W. W. Chu, editor, *Data Mining and Knowledge Discovery for Big Data*, number 1 in Studies in Big Data, pages 83–116. Springer Berlin Heidelberg, 2014. DOI: 10.1007/978-3-642-40837-3_3.
- [18] N. C. for Environmental Information. Earth observation. <http://ngdc.noaa.gov/eog/>. Accessed: 2016-06-22.
- [19] T. Gutierrez, G. Krings, and V. D. Blondel.
- [20] S. I. Hay, A. M. Noor, A. Nelson, and A. J. Tatem. The accuracy of human population maps for public health application. *Tropical medicine & international health : TM & IH*, 10(10):1073–1086, Oct. 2005.
- [21] J. V. Henderson, A. Storeygard, and D. N. Weil. Measuring economic growth from outer space. *The American Economic Review*, 102(2):994–1028, 2012.
- [22] ICF International. Dhs overview, 2016.
- [23] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman. Reading text in the wild with convolutional neural networks. *CoRR*, abs/1412.1842, 2014.
- [24] D. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, Dec. 2014. arXiv: 1412.6980.
- [25] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [26] O. R. N. Laboratory. Landscan. <http://web.ornl.gov/sci/landscan/>. Accessed: 2016-06-22.
- [27] C. Linard, V. A. Alegana, A. M. Noor, R. W. Snow, and A. J. Tatem. A high resolution spatial population database of Somalia for disease risk mapping. *International Journal of Health Geographics*, 9:45, Sept. 2010.
- [28] T. Louail, M. Lenormand, O. G. Cantú, M. Picornell, R. Herranz, E. Frias-Martinez, J. J. Ramasco, and M. Barthelemy. From mobile phone data to the spatial structure of cities. *arXiv:1401.4540 [physics]*, Jan.

Evaluating socio-economic state of a country analyzing airtime credit and mobile phone datasets. *arXiv:1309.4496 [physics]*, Sept. 2013. arXiv: 1309.4496.

2014. arXiv: 1401.4540.
- [29] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556 [cs]*, Sept. 2014. arXiv: 1409.1556.
- [30] F. R. Stevens, A. E. Gaughan, C. Linard, and A. J. Tatem. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *PLOS ONE*, 10(2):e0107042, Feb. 2015.
- [31] J. Stillwell and H. J. Scholten. *Land use simulation for Europe*, volume 63. Springer Science & Business Media, 2001.
- [32] A. J. Tatem, S. Adamo, N. Bharti, C. R. Burgert, M. Castro, A. Dorelien, G. Fink, C. Linard, M. John, L. Montana, et al. Mapping populations at risk: improving spatial demographic data for infectious disease modeling and metric derivation. *Population health metrics*, 10(1):1, 2012.
- [33] A. J. Tatem, N. Campiz, P. W. Gething, R. W. Snow, and C. Linard. The effects of spatial population dataset choice on estimates of population at risk of disease. *Population Health Metrics*, 9:4, 2011.
- [34] W. Tobler, U. Deichmann, J. Gottsegen, and K. Maloy. The global demography project (95-6). 1995.
- [35] W. R. Tobler. Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association*, 74(367):519–530, 1979.
- [36] A. Toshev and C. Szegedy. Deeppose: Human pose estimation via deep neural networks. *CoRR*, abs/1312.4659, 2013.
- [37] UNICEF. Manual for mapping and household listing, 2016.
- [38] UNICEF. The state of the world's children 2012: A fair chance for every child. Technical report, United Nations Children's Fund (UNICEF), 2016.
- [39] United Nations ChildrenâŽs Fund (UNICEF). Multiple indicator cluster survey, 2016.
- [40] United Nations Economic and Social Council. Resolution adopted by the Economic and Social Council on 10 June 2015. 020 World Population and Housing Census Programme, 2015.
- [41] United Nations Population Fund. Census reaches vulnerable women and girls in a remote area of myanmar for the very first time, 2016.
- [42] United Nations Statistics Division. Principles and Recommendations for Population and Housing Censuses. Revision 2. Technical Report No.67/Rev.2, United Nations Statistics Division, New York, 2008.
- [43] United States Geological Survey. What are the best spectral bands to use for my study?, 2016.
- [44] D. Vlahov, S. R. Agarwal, R. M. Buckley, W. T. Caiaffa, C. F. Corvalan, A. C. Ezeh, R. Finkelstein, S. Friel, T. Harpham, M. Hossain, et al. Roundtable on urban living environment research (ruler). *Journal of Urban Health*, 88(5):793–857, 2011.
- [45] Weiss, W. USAID. Personal communication to author, 2016-03-29.
- [46] WorldPop. Worldpop. <http://www.worldpop.org.uk/>. Accessed: 2016-06-22.
- [47] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon. Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping. *arXiv:1510.00098 [cs]*, Sept. 2015. arXiv: 1510.00098.
- [48] J. Zhou, H. Pei, and H. Wu. Early Warning of Human Crowds Based on Query Data from Baidu Map: Analysis Based on Shanghai Stampede. *arXiv:1603.06780 [cs]*, Mar. 2016. arXiv: 1603.06780.