Towards Sustainable Census Independent Population Estimation in Mozambique

Citation for published version:

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Towards Sustainable Census Independent Population Estimation in Mozambique

Isaac Neal$^1$, Sohan Seth$^1$\textsuperscript{*}, Gary Watmough$^1$, Mamadou Saliou Diallo$^2$
University of Edinburgh$^1$ and UNICEF$^2$
\{ineal, sseth, gary.watemough\}@ed.ac.uk, mamsdiallo@unicef.org

Abstract

Reliable and frequent population estimation is key for making policies around vaccination and planning infrastructure delivery. Since censuses lack the spatio-temporal resolution required for these tasks, census-independent approaches, using remote sensing and microcensus data, have become popular. We estimate intercensal population count in two pilot districts in Mozambique. To encourage sustainability, we assess the feasibility of using publicly available datasets to estimate population. We also explore transfer learning with existing annotated datasets for predicting building footprints, and training with additional ‘dot’ annotations from regions of interest to enhance these estimations. We observe that population predictions improve when using footprint area estimated with this approach versus only publicly available features.

1 Introduction

Accurate fine scale population estimates serve as a fundamental tool for policymakers. Many decisions involving access to services, distribution of vaccines and disaster relief, tracking of migration, and more are informed based on the most up to date population estimates for a region. Where these estimates are of insufficient resolution — either spatially or temporally — optimal decision making becomes difficult. Thus, there is a need for accurate and sustainable fine scale estimates of population globally, particularly in response to the COVID-19 pandemic, which requires efficient distribution of vaccines to vulnerable people, see e.g., [Wang et al., 2020].

Over time, census data loses its resolution, both temporal and spatial, making it difficult to inform decision making for many of the problems mentioned above. Censuses are conducted infrequently, typically decennially, and due to privacy constraints, census data is typically unavailable to the wider community at the smallest administrative levels (Wardrop et al., 2018). Furthermore, census data can be inaccurate or incomplete due to budget limitations, lack of training, and socio-political circumstances (Wardrop et al., 2015), and may quickly become outdated due to conflicts, rapid migration or urban development, and disasters (Engstrom et al., 2020).

Census-independent population estimation (or bottom-up estimation) uses updated demographic information in periodic household surveys, or microcensuses, and detailed visual information offered by remote sensing technology to predict intercensal population density in non-surveyed areas. This approach could improve both temporal and spatial resolution of census data, since remote sensing data is available on a regular basis with sufficient resolution to estimate population at a fine scale. A summary of recent works in this area is provided in Table 1.

In this paper, we present ongoing work on bottom-up population estimation in Mozambique with a focus on sustainability. The goal is to establish a pipeline that can be maintained and used by non-experts over a long period, that relies on existing tools and datasets as much as possible, and requires minimal human supervision, e.g., in terms of annotation. We assess the feasibility of this approach, and find that building footprint area is a crucial attribute for estimating population. Moreover, building footprint area estimates using existing deep architecture and transfer learning are improved if additional annotated data is used from the region of interest (ROI). For greater sustainability, we suggest using easier ‘dot’ annotation (Lempitsky & Zisserman, 2010) for buildings from the ROI rather than more time-consuming polygon annotation which requires more human supervision.
2 DATA AND PROCESSING

Microcensus: UNICEF funded a microcensus in 2019 conducted by SpaceSUR and GroundWork over two ROIs in Mozambique: Boanne (BOA) and Magude (MGD) (see Fig. 1A). The survey was conducted at a household level and households were exhaustively sampled over several primary sampling units (PSUs). We aggregated the household survey data to a 100 m grid to generate the response variable for predicting gridded population from remote sensing data (see Fig. 1B).

Remote Sensing: A variety of remote sensing imagery and derived data were used to model population count. The data sources are summarized in Table 2. Example images of each data source are shown in Fig. 1C. The 50 cm high-resolution satellite imagery (Vivid from Maxar) was a mosaic of WorldView-2 images, mostly from 2018 and 2019 (83% and 17% for BOA and 43% and 33% for MGD, remainder from 2011 to 2020). The night time lights (NTL) data was average intensity for the year 2019. The land cover classification was reclassified to condense forest types to open forest and closed forest. We ignored no class, closed forest, herbaceous wasteland, moss, bare, snow, water, and sea as they did not appear in our survey tiles. This left 5 classes. All raster data was reprojected to the 100 m survey grid for each ROI. Three types of resampling were used for different data-sources: average resampling for continuous data, nearest-neighbor resampling for coarser resolution categorical data, and any resampling (i.e. 1 if any underlying tile is 1 and otherwise 0) for finer resolution binary data.

Distance to Road Calculation: We rasterized the OSM road shapefiles to produce a Boolean raster indicating the presence of road in each grid tile, and calculated the Euclidean distance from each tile to the nearest tile with road as distance to road. We treated all road annotations from the OSM dataset the same, yielding one value for distance to road for each grid tile.

Building Footprint Identification We used a U-Net architecture (Ronneberger et al., 2015) with a VGG16 encoder (Simonyan & Zisserman, 2015) pre-trained on weights from ImageNet (Rusakovskiy et al., 2015) to segment an image into building and non-building areas. BFIa) We pre-trained the building footprint segmentation model on annotated satellite imagery from the SpaceNet challenge (Van Etten et al., 2019). We used the processed AOI 1 dataset from SpaceNet v1, which contains a large quantity of satellite imagery from Brazil with polygon building annotations. The dataset differs from our ROIs in Mozambique in several ways: SpaceNet tiles have more urban areas, a wetter climate, and different building colouration (see Figs. 1D-E). These factors hindered transfer learning between the tasks, and justified fine-tuning on a dataset from Mozambique. BFIb) We fine-tuned the model on ‘dot’ annotated buildings covering an area of 10 km² (i.e., 1,000 grid...
Table 1: Summary of recent literature on topic of bottom up population estimation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nigeria</td>
<td>0.5 m (Maxar)</td>
<td>10 m (Polygon data)</td>
<td>30 m (Landsat)</td>
<td>0.5m (Maxar), 100m (WorldPop), various (OSM school density, household size)</td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>12-30 m (Urban Footprint)</td>
<td>750 m (Night time Lights)</td>
<td>500 m</td>
<td></td>
</tr>
<tr>
<td>Bo, Sierra Leone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Input Resolution 90 m Village level City district level 100 m

Input Data Cost High (Maxar data) Free (public data) Free (public data) Free (OSM, WorldPop) High (Maxar data)

Performance Validation eTally data Train/test split LOOCV Train/test split

MEAPE 0.98 28 11 -

$R^2$ 0.58 0.26 -

NOTE: LOOCV = leave one out cross validation, MEAPE= median absolute percent error

Table 2: Remote sensing data sources, their characteristics and features extracted.

<table>
<thead>
<tr>
<th>Data</th>
<th>Resolution</th>
<th>Year</th>
<th>Frequency</th>
<th>Prerequisite</th>
<th>Publicly avail. (source)</th>
<th>Features Extracted (Count)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vivid Imagery</td>
<td>0.5 m</td>
<td>Various</td>
<td>4 d</td>
<td>N/A</td>
<td>No (Maxar)</td>
<td>Building area (1)</td>
</tr>
<tr>
<td>Landsat 8</td>
<td>30 m</td>
<td>2019</td>
<td>8 d</td>
<td>N/A</td>
<td>Yes (NASA/ USGS)</td>
<td>10 bands, NDVI, NDWI (12)</td>
</tr>
<tr>
<td>High Resolution</td>
<td>30 m</td>
<td>2015</td>
<td>N/A</td>
<td>High res. imagery and census data</td>
<td>Yes (CIESIN)</td>
<td>Binary Map (1)</td>
</tr>
<tr>
<td>Settlement Layer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land cover</td>
<td>100 m</td>
<td>2019</td>
<td>Annual</td>
<td>Various</td>
<td>Yes (ESA)</td>
<td>LCC (5)</td>
</tr>
<tr>
<td>classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Night-time lights</td>
<td>750 m</td>
<td>2019</td>
<td>1 d</td>
<td>N/A</td>
<td>Yes (NOAA)</td>
<td>Radiance (1)</td>
</tr>
<tr>
<td>Road Data</td>
<td>Vector</td>
<td>Various</td>
<td>N/A</td>
<td>Volunteer annotation</td>
<td>Yes (OSM)</td>
<td>Distance to road (1)</td>
</tr>
</tbody>
</table>

NOTE: Although WorldView-2 imagery may be available every 4 days, we only had access to Vivid data from Maxar collected from various years. The number in parenthesis in the final column shows the number of derived features from that dataset.

tiles to reduce manual annotation) from the ROIs (outside surveyed areas to avoid data leakage). These annotations are a subset of those produced through a mixture of automatic estimation and manual curation in previous work by SpaceSUR. The ‘dot’ annotation was converted to raster class labels by rasterizing a circle around each point with the average area of buildings in the respective ROI. Dot annotation was preferred over polygon annotations to avail sustainability since the former is less time-intensive than the latter. This approach might not infer the exact boundary of a building but is only an intermediate step to population estimation.

Context Variable We introduced the notion of context to our model by calculating the average of each extracted feature (except distance to road) in the tiles surrounding the tile of interest (two surrounding contexts, 8 tiles and 24 tiles), and using these values as additional features in the model. In doing so, we aimed to provide the model with an understanding of the features on coarser scales.

Non-representative Tiles Upon investigating the data, we concluded that, due to the retrospective nature of the data, several tiles at the boundary of the PSUs only covered a PSU partially, and several surveyed areas had developed either before or after our high resolution satellite imagery was captured. This meant that the tiles often contained non-surveyed buildings or missing buildings, and therefore, our gridded data contained both developed tiles (i.e., with a large number of buildings) labeled as low population, and undeveloped tiles labeled as high population. Although it is possible to either automatically detect these ‘outlier’ tiles during training or use a robust loss function that is insensitive to them, it is not possible to validate a model on non-representative data. Therefore, we manually excluded these grid tiles from the data by comparing the surveyed buildings (GPS locations available in the microcensus) with those appearing in the Vivid imagery.
3 Results and Discussion

**Dataset** We have 199 survey grid tiles and 61 predictor variables (see Table 2). **Cross-validation** Due to the limited quantity of data, we evaluated our population model through 4-fold cross validation. The data is split spatially into four approximately equal sized subsets (for each ROI separately), and we reported the error metrics over pooled prediction from the four validation folds. **Model** Due to the limited availability of data, we fit a single linear model with Huber loss function and $\ell_1$ regularization (Yi & Huang, 2017) for both districts to model the log of population count given a combination of public features and BFIa or BFib. We chose the model hyperparameters using 3-fold cross-validation over the training set for each fold of our spatial cross validation. **Evaluation Metrics** We chose several evaluation metrics, i.e., $R^2 = 1 - \sum_i (y_i - \hat{y}_i)^2 / \sum_i (y_i - \bar{y})^2$, median absolute error $\text{MAE} = \text{median} |y_i - \hat{y}_i|$, median absolute percentage error $\text{MAPE} = \text{median} |y_i - \hat{y}_i| / y_i$, adjusted median absolute percentage error $\text{AMAEPE} = \text{median} |y_i - \hat{y}_i| / (y_i + 10)$ (to avoid division by zero), and aggregated percentage error $\text{AGGPE}(A) = \sum_{i \notin A} |y_i - \sum_{i \notin A} \hat{y}_i| / \sum_{i \notin A} y_i$ (to capture error at a ROI level). **Null model** The null model predicted the population as the mean of the training set irrespective of the feature values. **Results** We observe that the model can effectively predict population, and outperforms the null model. The model performs the best with either public and fine-tuned building footprints (BFib) as features, or only BFib as features, and the performances are similar. A loss in accuracy is incurred when using either public only or public and pre-trained building footprints (BFia) as features, however, the accuracy is still acceptable (MAE is 3.8 versus 7.5 in null model). The model performed the worst when using only pre-trained building footprints (BFia) as features, indicating that fine-tuning improves performance substantially.

<table>
<thead>
<tr>
<th>Features used</th>
<th>$R^2$</th>
<th>MAE</th>
<th>AMAEPE</th>
<th>MAPE</th>
<th>AGGPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>0.05</td>
<td>51.8</td>
<td>0.23</td>
<td>3.84</td>
<td>25.4</td>
</tr>
<tr>
<td>BFia</td>
<td>-0.08</td>
<td>59.9</td>
<td>0.25</td>
<td>4.02</td>
<td>32.1</td>
</tr>
<tr>
<td>BFib</td>
<td>0.54</td>
<td>39.2</td>
<td>0.20</td>
<td>3.41</td>
<td>14.9</td>
</tr>
<tr>
<td>Public + BFia</td>
<td>0.05</td>
<td>50.1</td>
<td>0.23</td>
<td>3.97</td>
<td>27.3</td>
</tr>
<tr>
<td>Public + BFib</td>
<td>0.53</td>
<td>42.1</td>
<td>0.19</td>
<td>3.45</td>
<td>13.2</td>
</tr>
<tr>
<td>Null Model</td>
<td>-0.12</td>
<td>76.4</td>
<td>0.41</td>
<td>7.57</td>
<td>1.68</td>
</tr>
</tbody>
</table>

See Evaluation Metrics above for metric definitions. BFia and BFib are pre-trained and fine-tuned building area estimates respectively. Predicted vs. observed plot (right) summarizes the results for Public + BFib.

We are currently addressing the following improvements to the model: **Buildings under construction** Our building segmentation model often identifies buildings that are under construction or buildings that are in ruin (e.g., historical buildings from colonial era with roof caved in). Since building footprint is an important feature in predicting population this can introduce bias in the outcome. We are improving the building segmentation model to help reduce this bias. **Non-residential buildings** Our building segmentation model does not discriminate between non-residential buildings (e.g., grain storage or schools) and residential buildings. We are addressing this by computing individual footprints prior to total area and discounting buildings that are too large or too small to likely be inhabited. **Distance to road** We treat all road types to be the same. We are extending this approach to give a more informative set of distance variables (e.g. distance to major highway) for different road types. We also intend to calculate road density to compare usefulness. **Sustainability** Sustainability is an issue in most bottom-up estimation techniques in the literature. The causes for this are twofold: the high cost of conducting population surveys, and limitations inherent to the machine learning methodology that might require expert annotation. Our results show that population can be estimated with limited survey data and ‘dot’ annotations. We are addressing this further by exploring ways to reduce dot annotation required. **External validation and transferability** It is difficult to externally validate the existing population estimation approaches or to compare them. This is because existing methods have been applied on different countries and a standard microcensus dataset does not exist for comparison. Furthermore, when microcensus data does exist it is difficult to access satellite images with the same spatial, temporal and radiometric characteristics using the Vivid data product. We are addressing this through conducting fresh microcensus surveys in Mozambique.
ACKNOWLEDGMENTS

The project was funded by the Data for Children Collaborative with UNICEF. The authors thank Pablo Martín Amieva (SpaceSUR) for his valuable insights on the project.

REFERENCES


