

Exploring the automatic generation of census pre-enumeration areas (preEA tool)

Sarchil H. Qader¹, Andrew Harfoot², Mathias Kuepie³, Sabrina Juran⁴, Rosa Catalina Valle Piñuela⁴, Attila Lazar¹, Andrew J Tatem¹

¹WorldPop, Department of Geography and Environment, University of Southampton, Southampton, UK

²GeoData, Department of Geography and Environment, University of Southampton, Southampton, UK

³United Nations Population Fund, New York

⁴United Nations Population Fund, Regional Office for Latin America and the Caribbean, Republic of Panama

Short abstract

Enumeration Areas (EAs) are the operational geographic units for collecting and disseminating census data and are often used as a national sampling frame for various types of surveys. In certain low income or conflict-affected countries, EA demarcations are incomplete, outdated, or missing. Even for stable and middle or high-income countries, creating and updating EAs is a challenging yet essential task in preparing for a national census. Commonly, EAs are created by manually digitising small geographic units on high-resolution satellite imagery or physically walking the boundaries of units, both of which are time, cost, and labour intensive. We have developed a user-friendly tool that could be employed to generate draft EA boundaries automatically. The tool is based on high-resolution gridded population and settlement datasets, GPS household locations, building footprints and uses publicly available natural, man-made and administrative boundaries. Initial outputs were produced in Burkina Faso, Somalia, Paraguay, Togo, Zimbabwe, Niger and Guinea. The results indicate that the EAs are in line with international standards including boundaries that are following ground features and easily identifiable, no overlaps, the boundaries are nested within administrative boundaries, and they are compact and free of pockets and disjoints.

Extended abstract

Historically, the creation of census Enumeration Areas (EAs) has been an expensive task solved by approaches such as physically walking to map EA boundaries, which can require intensive resources and multiple years to complete (Lu, 2009; Yacyshyn and Swanson, 2011). For instance, in Zambia, the 2010 census mapping exercise was expected to take about two years to be completed at a total cost of US \$7 million (Yacyshyn and Swanson, 2011). Since the advent of Geographic Information Systems (GIS) and high-resolution satellite photography, census cartographers in some countries have been able to manually digitise EA boundaries from satellite imagery, leading to better control of EA delineation. However, manual digitisation of EAs is still costly and labour-intensive. It is also prone to human error and can have poor accuracy (Alazab et al., 2009; Balinski et al., 2010; Cockings et al., 2011). In addition, when manually digitising EAs, it can be challenging to accommodate population and area constraints. Furthermore, EAs or sampling frames in countries with large population displacements and rapid urbanisation need regular updating, meaning that the manual digitisation process would need to be replicated. Combining automation and manual approaches could be a viable middle ground to accelerate this process.

This work represents the first attempt to generate an automatic census pre-enumeration area and a population sampling frame where the boundaries are based on vector data for roads, settlements and other physical features. The census pre-Enumeration Area (PreEA) tool has been developed as part of Geo-Referenced Infrastructure and Demographic Data for Development (GRID3) being led by WorldPop and in close collaboration with United Nation Population Fund (UNFPA). It is an automated zone design tool intended to generate census pre-EAs. The tool has been developed in the Python programming language and it has been converted to a user-friendly tool in the QGIS platform. The preEA tool operates by subdividing the target area into as small as possible regions using existing digital boundaries - such as roads, rivers and governmental jurisdictions (See figure 1 for the input datasets in Burkina Faso). The user specifies parameters that influence how these very small regions are merged together, and the tool continues merging until all the constraints are met. This creates a frame with complete non-overlapping coverage that follows logical boundaries while meeting local criteria for EA characteristics (See the outputs of the tool in figure 2 and 3). In the context of EA delineation, the tool is designed to meet a range of applications to assist with EA creation. In countries where digitized EA already exist, the tool can keep the old EA boundaries and modify these EAs that, due to demographic changes, are now estimated to exceed the maximum household size. In countries where digitized EAs are non-existent, the tool can be used to create a new set of EAs that meet the country's need in terms of area, population size and other constraints.

It is important note that the suggested outcomes of the application process of the pre-EA tool needs to be validated on the ground during the cartographic exercise and the household listing prior to the actual census unless insecurity restricts access.

Even though the primary purpose of this toolset is the automatic delineation of census pre-EAs, it can also be used to define sampling frames in the interim for socio-economic household surveys. Household surveys are the main source of demographic, health, and economic data in low- and middle-income countries. For designing these surveys, an up-to-date population count, an EA map and a sampling frame are necessary. However, due to political sensitivity, security and lack of resources, these datasets are frequently outdated and/or inaccessible. While waiting for updated census data and census EAs, PreEA tool can be used to overcome these issues by creating a sampling frame based on user requirements (population and size of the sampling frame) while respecting ground boundaries.

PreEA tool outputs

The preEA tool has been piloted in several countries and the outputs were assessed and validated by National Statistics Organizations (NSOs) using different approaches. For instance, the Bureau Central de Recensement (BCR) in the Democratic Republic of Congo has carried out a series of field tests in a range of urban and rural contexts to assess the suitability of the automatic approach to creating preEA boundaries. The fieldwork was successfully conducted, and expectations were met and even exceeded. The preEA outputs were found to help facilitate enumeration, as the BCR teams could navigate within the pre-EA boundaries with ease and count every person within these areas. Recently, Zimbabwe National Statistics Agency (ZIMSTAT) has compared the automatic preEA outputs in 10 wards to a recent socio-economic survey. The team indicated that insignificant differences were found between the two products. In other countries such as Togo, Somalia, Paraguay, Guinea, Burkina Faso and Niger, NSOs have adopted other techniques to evaluate the preEA outputs including integration with high resolution satellite imagery, local knowledge, old census data, building footprints and recently collected population data for manual revision. Overall, the results were promising, and the

EA criteria were ensured to be in line with international principles and recommendations including, i) EA boundaries are mutually exclusive (non-overlapping) and exhaustive (cover the entire country), ii) The outlines of the EA boundaries are in line with ground feature boundaries and easily identifiable in the urban and populated areas, iii) All the generated EA boundaries are nested within administrative boundaries and they are not crossing through buildings, iv) The EAs are compact and free of pockets or disjoint sections, v) The population size and area were set to thresholds such that each EA can be small enough and accessible to be covered by an enumerator within the census period, vi) In addition, with respect to the population and area constraints, the created EAs are large enough to guarantee data privacy, vii) Furthermore, the EAs are flexible enough to allow the widest range of tabulations for different statistical reporting units. The sampling frame derived from our EA delineation process is reproducible and complete, making it for other data collection activities such as nationally representative household surveys, which typically rely on census data as a sampling frame.

To provide more details on the application of the tool, the initial inputs and outputs of the tool for Burkina Faso are shown in Figure 1, 2 and 3. To generate the census pre-EAs in Burkina Faso, the maximum population size was set to 1,000 people and the geographic area to 9 km². In some areas, the maximum thresholds have been exceeded where regions have not been sufficiently subdivided by the digital boundaries (Figure 2b). This may be due to insufficient detail being present in the digital boundaries or it may be deliberate to prevent settlements from being subdivided. In other areas, EAs have low population values (Figure 2b) because either the geographic area constraint has been reached, or further merging was prevented by the presence of uncrossable borders such as rivers, main roads or administrative boundaries. Figure 3 shows that the automatic census pre-EA outputs are in line with international principles.

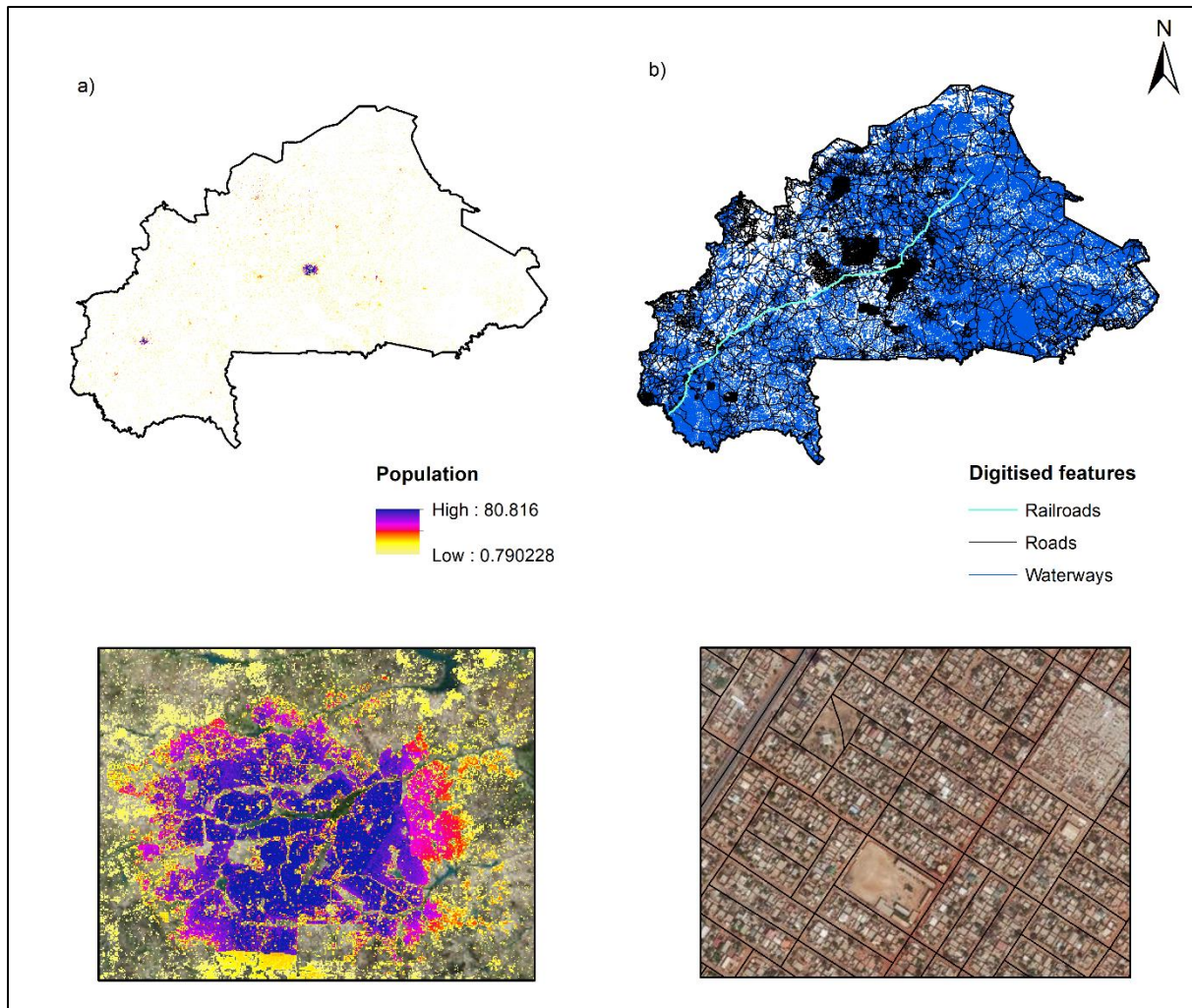


Figure 1. Input datasets to the tool for creating the census pre-Enumeration Areas in Burkina Faso; a) represents the high-resolution gridded population estimates (100mx100m) (WorldPop and Institut National de la Statistique et de la Démographie du Burkina Faso. 2020), b) represents the digitized georeferenced layers such as roads, railroads and waterways (OSM, 2015).

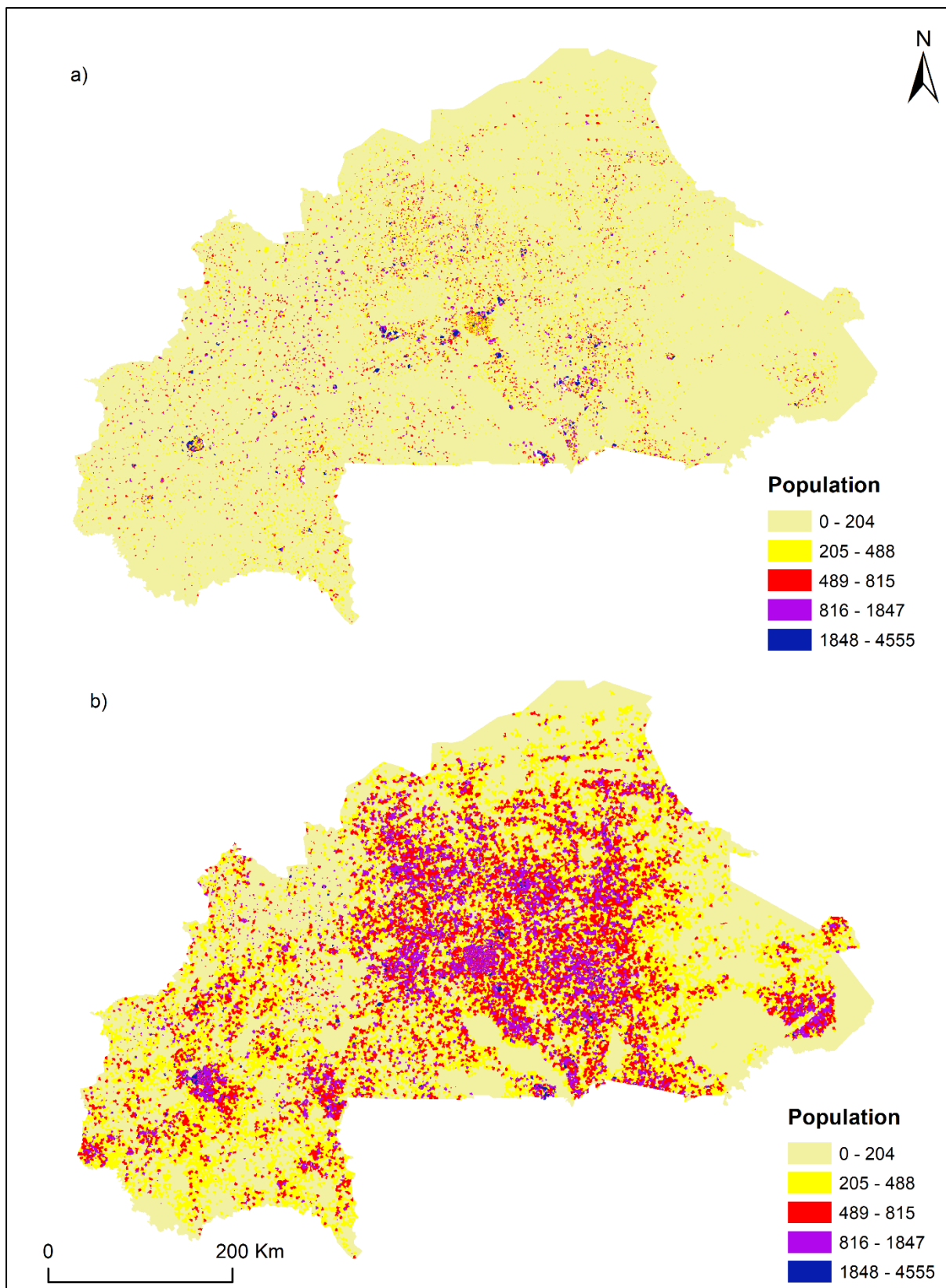


Figure 2. Shows the preEA tool's outputs; a) shows the building block polygons: these are the polygons generated from the input data, with population assigned but prior to any merging taking place, b) generated candidate EAs: the polygons generated following the completion of the merging process. After the merging process, some EAs might have populations larger than the maximum thresholds

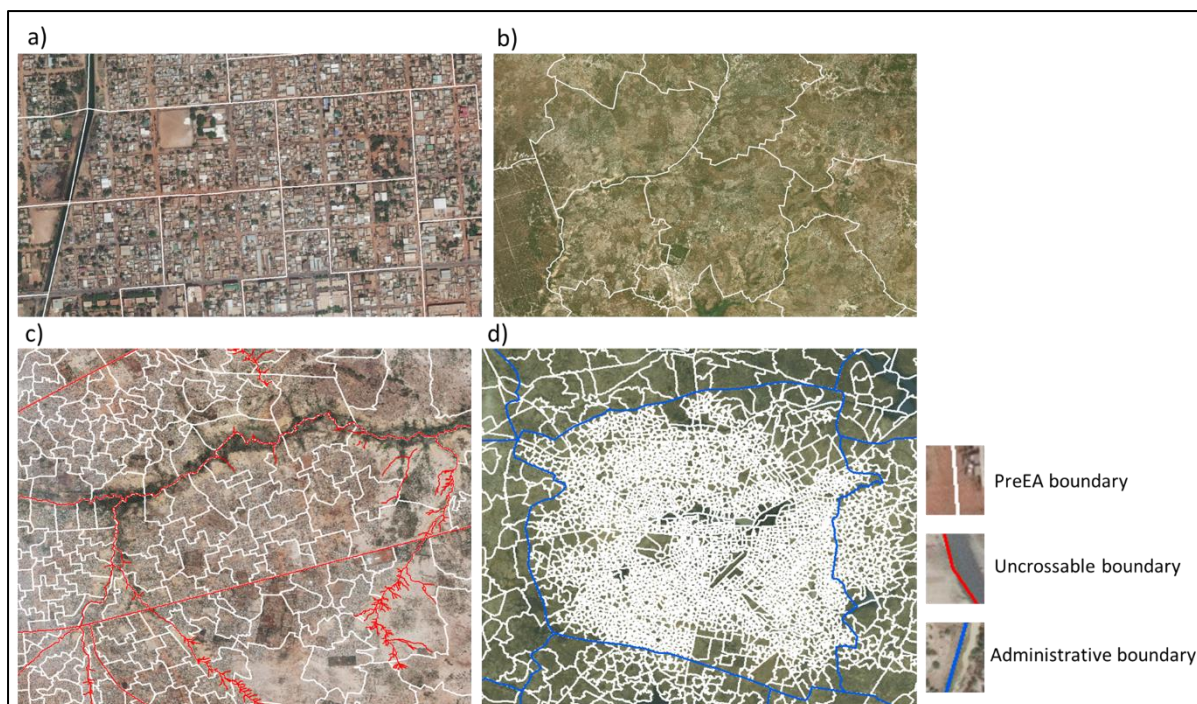


Figure 3. Outline of pre-census EAs in different locations over Burkina Faso generated by preEA tool. In the given examples, (a) represents urban examples, (b) represents rural examples, (c) shows the EAs are respecting the uncrossable features, and (d) shows the EAs are nested within administrative boundaries.

Conclusions and policy implications

This collaborative work represents the first attempt to generate automatic outputs of census pre-enumeration areas and a population sampling frame where the boundaries are based on vector data for roads, settlements and other physical features. The process of EA delineation is a crucial early step in census preparation as it helps in budgeting and allocating materials, logistical and personnel requirements, and is a key factor in the success of the census. Our algorithm simplifies the complex and time-consuming processes of EA delineation that is crucial in the preparation of the national census.

The preEA tool will be introduced to National Statistical Offices and relevant organization as a means to adopt a cost-effective tool to help them with their EA delineation work.

Reference

- Yacyshyn AM, Swanson DA (2011) The Costs of Conducting a National Census: Rationale for Re-Designing Current Census Methodology in Canada and the United States. The 21st Annual Warren E. Kalbach Population Conference, at the University of Alberta, Edmonton, Canada
- Alazab M., Islam M., Venkatraman S. (2009) Towards Automatic Image Segmentation Using Optimised Region Growing Technique. In: Nicholson A., Li X. (eds) AI 2009: Advances in Artificial Intelligence. AI 2009. Lecture Notes in Computer Science, vol 5866. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-10439-8_14

Balinski M, Johnston R, McLean I, Young P (2010) Drawing a new constituency map for the united kingdom: the parliamentary voting system and constituencies bill 2010. The British Academy, London

Cockings S, Harfoot A, Martin D, Hornby D (2011) Maintaining existing zoning systems using automated zone-design techniques: methods for creating the 2011 Census output geographies for England and Wales. *Environ Plan A* 43 (10):2399–2418.
<https://doi.org/10.1068/a43601>

OpenStreetMap contributors. (2015) Planet dump [Data file from 2019]. Retrieved from <https://planet.openstreetmap.org>

WorldPop and Institut National de la Statistique et de la Démographie du Burkina Faso. 2020. Census-based gridded population estimates for Burkina Faso (2019), version 1.0. WorldPop, University of Southampton. doi: 10.5258/SOTON/WP00687.