Automatic national preEnumeration Areas (preEAs) for Burkina Faso (2019)

version 1.0

WorldPop, University of Southampton

26/01/2022

Prepared by: Sarchil Qader

Contents

1. Introduction

GRID3 is supporting the realisation of a fully digital census in Burkina Faso (BFA). BFA has conducted the fifth national population and housing census in December 2019-Janaury 2020. For the purpose of the census data collection, the territory of Burkina Faso was divided into operational geographic units called Enumeration Zones (Zones de Dénombrements (ZDs)) which is equivalent to Enumeration Areas (EAs). The main limitation of this census is that the EA boundaries have been drawn manually on paper and there is currently no national digitized EA dataset for BFA. This only allows the data analyses and reporting in table format, and thus makes dissemination and analysis of census data less informative and accessible. In addition, since the current EA boundaries are not georeferenced, accurate statistical summarization of the collected georeferenced census data at the EA level or for custom areas and representation of these on digital maps is a challenge. Finally, it can cause uncertainties and issues during household surveys whether the accurate area is surveyed. GRID3 supports the National Institute of Statistics and Demography (INSD) in digitalisation by semi-automatically creating preEnumeration Area datasets that can speed up and make the digital EAs demarcation more robust.

Commonly, EAs were done by manual digitization of small geographic units on high resolution satellite imagery. If such a technique is adopted alone, the process of EA delineation is highly time, cost and labour intensive. To respond to this challenge, WorldPop at the University of Southampton and UNFPA have developed a semi-automatic designation tool for delineating preEnumeration Areas (preEAs) and national population sampling frames in the absence of recent census data. This approach is based on high-resolution gridded population and settlement datasets and can use publicly available natural and administrative boundaries if reliable and detailed dataset do not exist in country. The approach can be used to create preEAs in countries where the EAs are non-existent or to update existing digitized EAs in countries where the EAs are outdated. It should be noted that the suggested EA approach outcomes need to be reviewed in the lab and manually edited as needed prior to the fieldwork. Then, they have to be validated on the ground during the cartographic exercise and the household listing prior to the actual census unless insecurity restricts access.

Purpose:

This report provides details on the data inputs and steps to generate national preEnumeration Areas (preEAs) in Burkina Faso (BFA).

Location:

- Burkina Faso

2. Methodology

2.1 Data Inputs

Name	Data Type	Date (receivin	Source	Description

		g the data)		
ADM commune boundaries	Polygo n	06/2020	Geographi c Institute of Burkina Faso (IGB)	Administrativ e Commune boundary (351 Communes)
HYD Course eau	Line	06/2020	IGB	Digitized waterways
Voie Railroads		06/2020	IGB	Digitized rail roads
OSM Roads	Line	10/2020	Open Street Map (OSM)	Digitized road types with road type labels
OSM Waterways	Line	10/2020	OSM	Digitized main waterways
BFA_population_v1_0_gridded.t if	Raster	10/2020	WorldPop	High-resolution gridded population dataset (100m x 100m)

2.2 Data preparation

This step includes the collation of accessible datasets (Section 1) and GIS processing of these to check and correct geometry (as needed) to be able to use them as inputs in the preEA tool.

2.3 Population raster

The main data inputs for this approach include GRID3 (version 1.0) high-resolution gridded population estimates at approximately 100m spatial resolution, produced for Burkina Faso (WorldPop and Institut National de la Statistique et de la Démographie du Burkina Faso. 2020). WorldPop disaggregated the published admin3 level census totals (Institut National de la Statistique et de la Démographie, 2019) into high resolution gridded estimates. Thus, the population raster represents the official and latest population numbers.

2.3.1 Geographic feature datasets

The line datasets were obtained from the Geographic Institute of Burkina Faso (IGB) and Open Street Map (OSM) (OSM, 2015) and the following preprocessing were applied to the datasets using ArcGIS software:

1- Clip all data lines to the official BFA country boundary
2- Re-project the input data to WGS 1984 UTM Zone 30N (Figure 1).

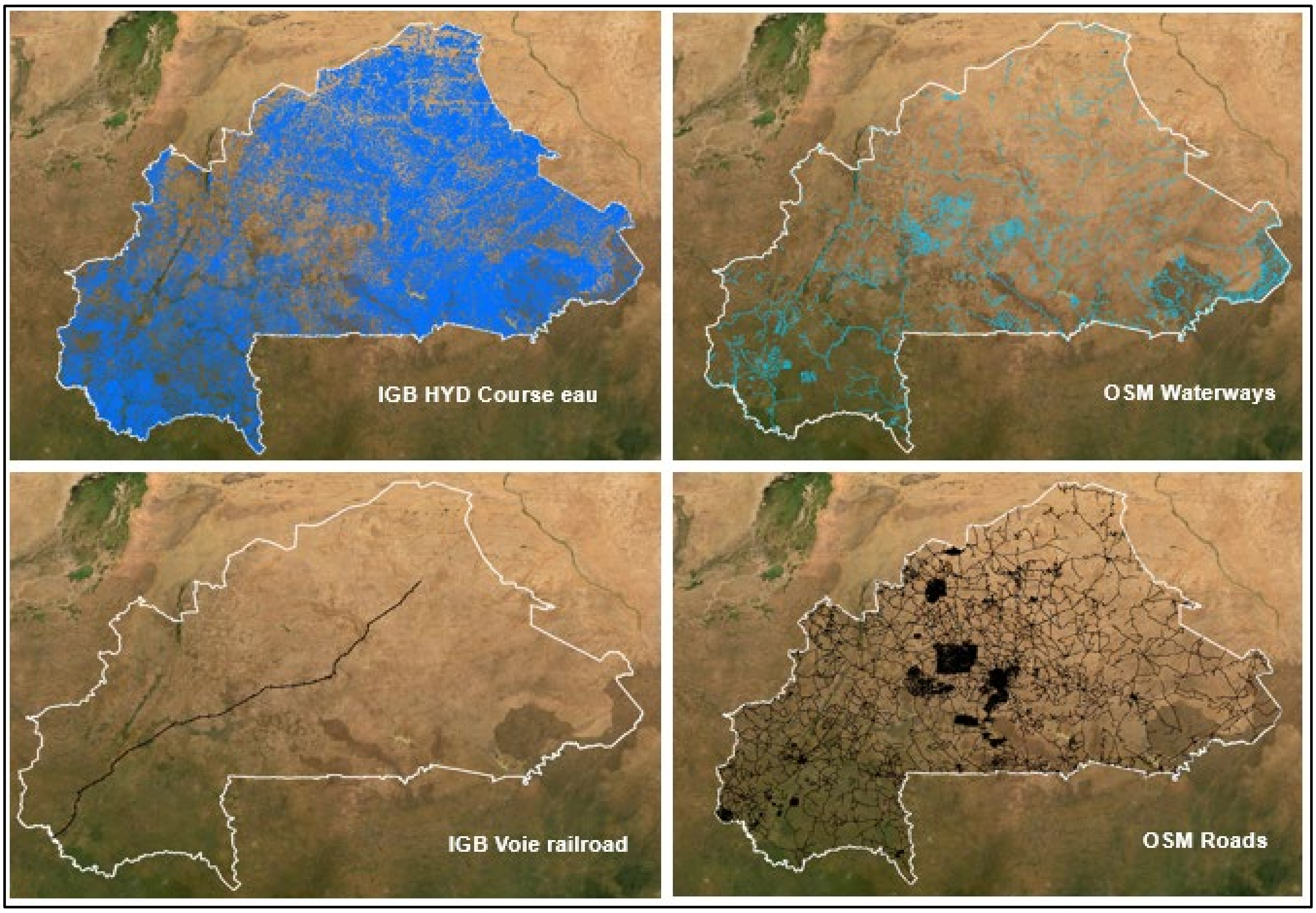


Figure 1. Shows the input line datasets

3- Generate single line road features in place of matched pairs of divided road lanes (to solve issues of the double or triple parallel highway or motor way lines) (Figure 2)

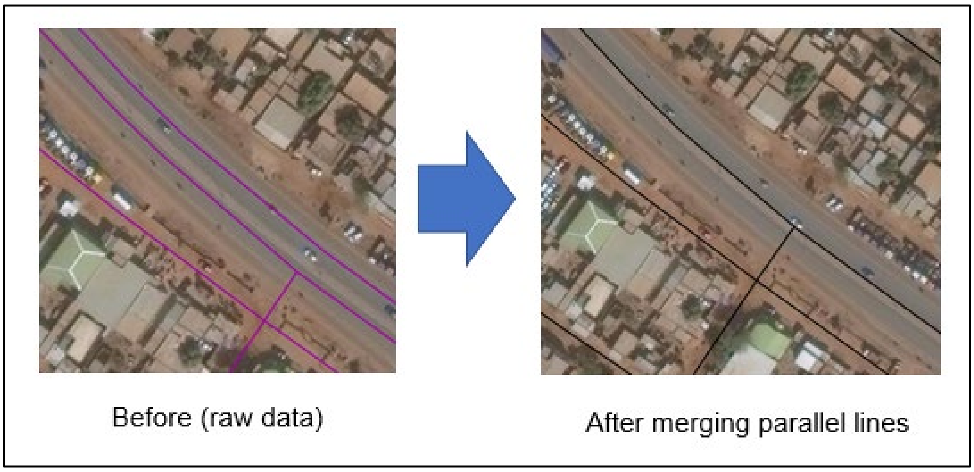


Figure 2. Shows the output of the road dataset after merging parallel lines such as motor ways.

4- It is important for EAs to not cross the main road as much as possible, therefore, specific road types were selected and set as uncrossable features for this work. In the OSM road dataset, the road types with fclass primary, trunk and tertiary were extracted and set as uncrossable feature lines. In addition, the OSM waterways (fclass= river) were also considered as uncrossable lines and it was merged with the main road types.
5- In areas where identical or very close water lines exist in both OSM waterways and IGB Hyd Course eau datasets, only lines from OSM waterways were retained. The corresponding lines in the IGB Hyd Course eau dataset were removed. This step was implemented to avoid creating silver polygons which may cause issues in the merging process.

2.4 Settlement delineation

In order to create boundaries around each settlement and minimize splitting small settlements such as villages into different preEAs (unless they are larger than the given maximum population and area constraints), the following steps were undertaken using ArcGIS software:

1- Create a raster file showing the settled pixels, where the pixel value of 1 is settled and NA is not settled. This is achieved by changed all >0 values of the population raster to 1.
2- Convert this binary settled area raster dataset to polygon features. This step identifies the outline of the settled areas.
3- The minimum bounding geometry was applied to create a feature class containing polygons that represent a specified minimum bounding geometry enclosing each input feature or each group of input features. In the geometry type options, the convex hull technique was selected.

4- The minimum bounding geometry does not respect neighbouring boundaries and results in creating a lot of overlaps. Therefore, an aggregation method was applied to combine polygons within 100 m distance of each other into new polygons. Setting the aggregation distance to 100 m is to make sure that the close settled areas are combined into one polygon.
5- To avoid creating complex shapes, step 3 was applied again to create a convex hull around each newly created polygon. However, this may generate an overlap among the convex hull boundaries again. Therefore, the aggregate polygon technique (step 4) was employed again but this time with only a 2 m distance.
6- Apply “multipart to singlepart” features to assure that multipart features are separated.
7- The total population of each convex hull polygon was then computed using a high gridded population dataset.
8- To avoid splitting larger settled areas particularly in rural strata into multiple preEAs, we used an observation-based threshold to classify the convex hull polygons into two categories: settlements containing (i) 650+ people or (ii) less than 650 people (Figure 4). Settlements with population equal and higher than 650 are relatively big and their boundary should be kept in the splitting process. This step will help keep relatively large settled areas such as villages in one preEA or more if the total population exceeded the maximum threshold, but they will be nested within the settlement extent (convex hull boundary).

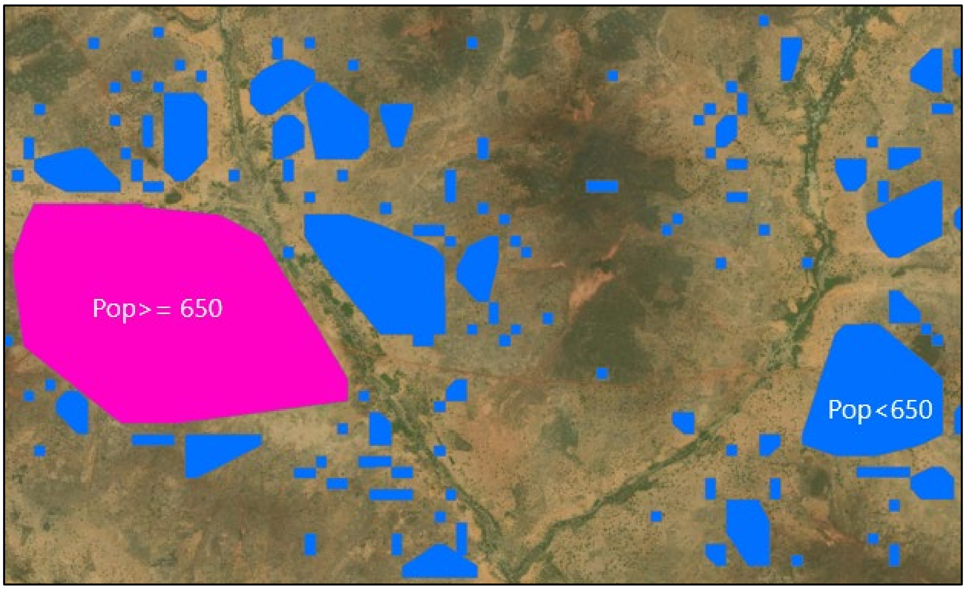


Figure 3. Shows the convex hull polygons with their total population. Pink colour represents polygons with 650 population and higher while blue colour represents polygons containing less than 650 people.

9- It is apparent from figure (3) that the polygons generated from the convex hull technique are not adjacent. Therefore, we have employed another technique to create adjacent polygons around each settled area (convex hull polygons) to assure a completeness across the country. Euclidean Allocation was adopted since it can produce adjacent polygons across the area of interest (Figure 4). Euclidean Allocation computes, for each cell, the nearest source (settled area) based on Euclidean distance. But here we are not interested in distance calculation, Euclidean Allocation was created to define and delineate proximal regions around individual data polygons by using polygon boundaries. The input data to the Euclidean allocation technique was convex hull polygons that have less than 650 people. Following this step adjacent polygons were created for the entire area and these adjacent polygons are proximal zones around convex hull outlines (settled areas) without cutting them.

10-The Euclidean Allocation output is a raster dataset; thus, it was converted to polygon features then clipped to the BFA official boundary.

11-The above-described Euclidean Allocation-based calculation produced adjacent polygons across the entire country including the settled areas where the population is larger than 650 people. However, we only needed this technique to generate adjacent polygons for areas where the total population for the convex hull polygons were less than 650 people. Therefore, the convex hull polygons with 650 people and higher (output of step 8) were used to remove corresponding polygons in the Euclidean allocation output and thus retain the original larger polygons.

12-The adjacent polygons generated from the Euclidean Allocation technique (Step 12) and convex hull polygons containing 650 people and higher were merged. Following this step, adjacent polygons were generated for entire BFA (Figure 4).

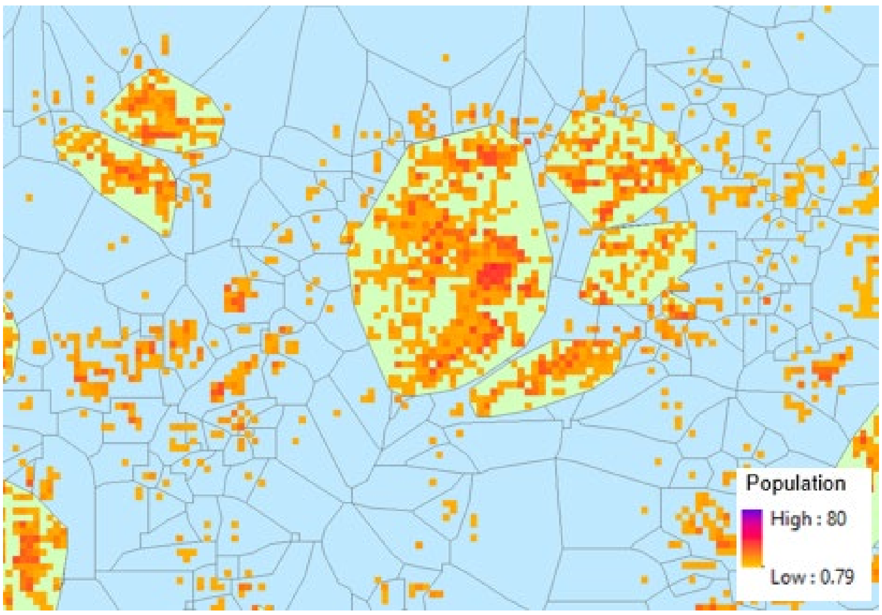


Figure 4. Example area showing full coverage adjacent polygons for BFA. The green colour represents polygons generated from convex hull techniques with a total population of 650 and higher. The blue colour represents the adjacent polygons generated from the Euclidean Allocation technique with a total population lower than 650.

13- The convex hull polygons containing 650 and higher were converted to line features and merged with the main road type lines and final waterways to build the final uncrossable feature input of the preEA tool.

2.5 Implementing the preEA tool

The prepared datasets in *Section 1* were entered as inputs into the preEA tool (unpublished) to conduct the automatic splitting and merging process in BFA. Using gridded population data and digitised geographic features, including administrative boundaries, natural and man-made features, the tool divided the country into small areas (building blocks) which are then merged to meet the criteria specified by the user in terms of population, geographic area and any other constraints. The tool was developed as a form QGIS plugin and implemented in QGIS. The tool's interface has two three main tabs (Figure 5): The first tab allows the user to input the necessary datasets, the second tab is dedicated for the merging criteria and tuning the parameters and the last tab is for the output names. In each tab there is a help section which provides summary guidance on each input, parameter and output.

The BFA preEA boundaries were created using the following settings:

Tool Input: Polygon and line datasets to generate polygons from

- Dataset 1: OSM road
- Dataset 2: OSM Waterways
- Dataset 3: IGB HYD Course eau
- Dataset 4: IGB Voie railroad
- Dataset 5: a merged dataset from Euclidean Allocation output and convex hull polygons containing 650 people and higher (Section 2.4, step 11).

Tool Input: Population raster

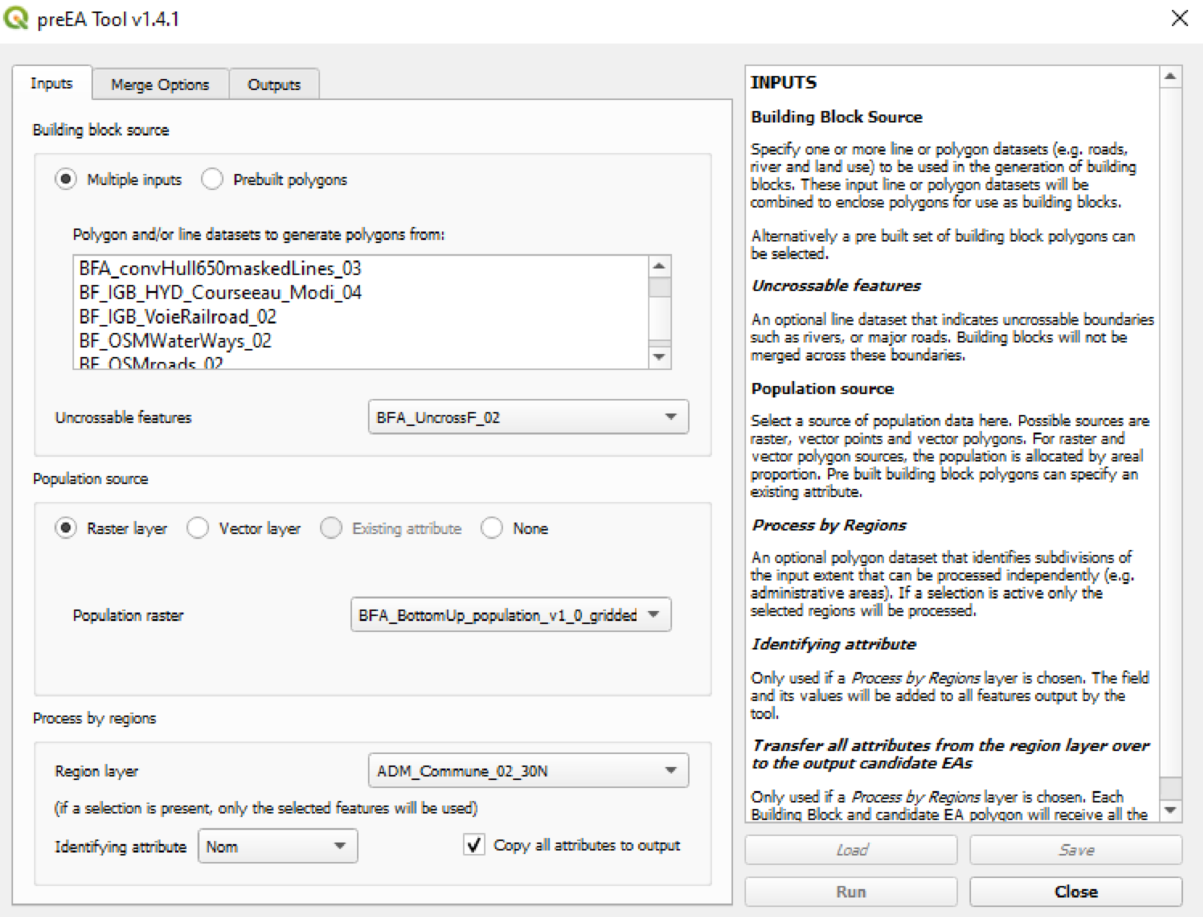
Dataset: BFA_population_v1_0_gridded.tif

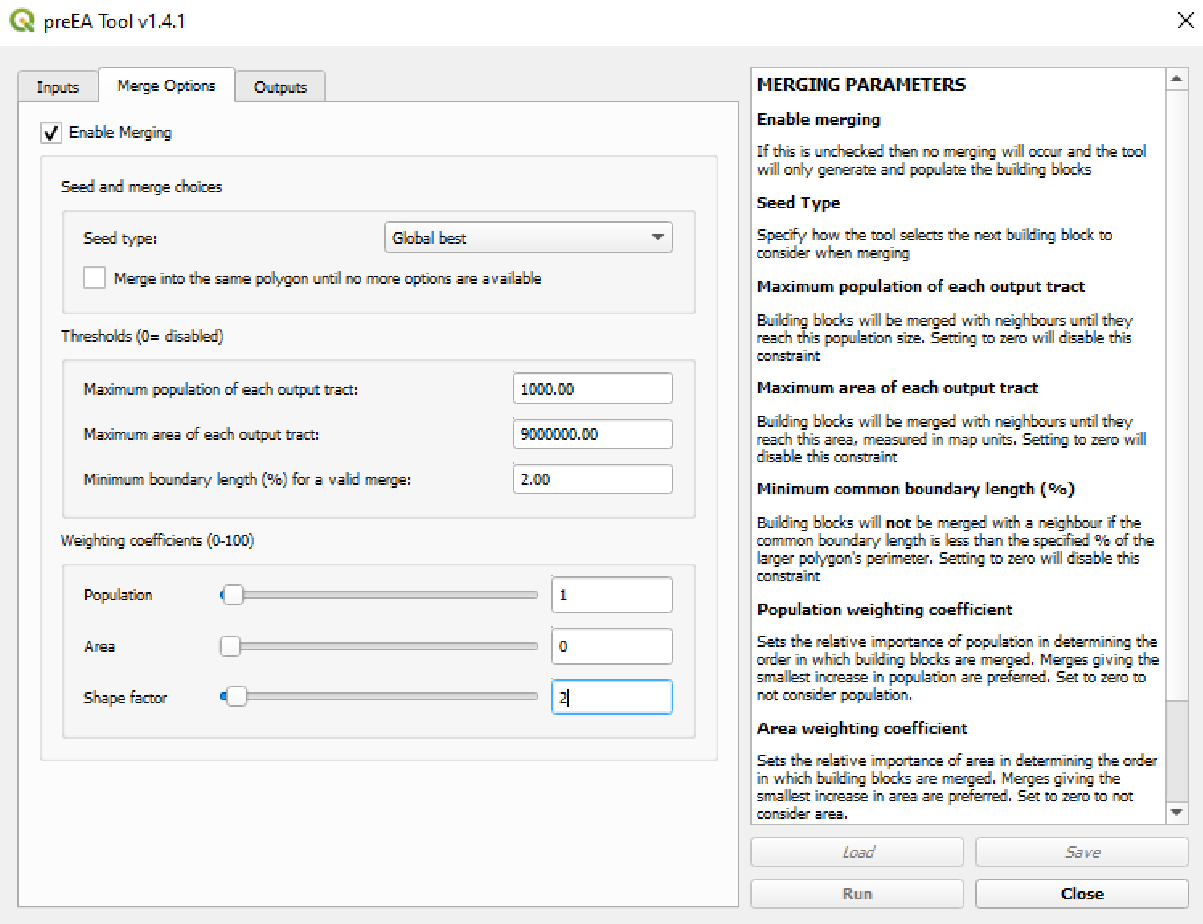
Tool Input: Uncrossable features [optional]

- Dataset: Merged dataset from the specific main roads, OSM waterways and convex hull lines containing 650 people and higher (Section 2.4, step 12).

For the rest of setting, please see figure 5.

Figure 5 shows the specific setting that were made in the preEA tool to generate automatic preEAs for Burkina Faso.


preEA Tool v1.4.1
Inputs
Merge Options
Outputs
Building block source
Multiple inputs
Prebuilt polygons
Polygon and/or line datasets to generate polygons from:
BFA_convHull650maskedLines_03
BF_IGB_HYD_Courseeau_Modi_04
BF_IGB_VoieRailroad_02
BF_OSMWaterWays_02
Uncrossable features
BFA_UncrossF_02
Population source
Raster layer
Vector layer
Existing attribute
None
Population raster
BFA_BottomUp_population_v1_0_gridded
Process by regions
Region layer
ADM_Commune_02_30N
(if a selection is present, only the selected features will be used)
Identifying attribute
Nom
Copy all attributes to output
INPUTS
Building Block Source
Specify one or more line or polygon datasets (e.g. roads, river and land use) to be used in the generation of building blocks. These input line or polygon datasets will be combined to enclose polygons for use as building blocks.
Alternatively a pre built set of building block polygons can be selected.
Uncrossable features
An optional line dataset that indicates uncrossable boundaries such as rivers, or major roads. Building blocks will not be merged across these boundaries.
Population source
Select a source of population data here. Possible sources are raster, vector points and vector polygons. For raster and vector polygon sources, the population is allocated by areal proportion. Pre built building block polygons can specify an existing attribute.
Process by Regions
An optional polygon dataset that identifies subdivisions of the input extent that can be processed independently (e.g. administrative areas). If a selection is active only the selected regions will be processed.
Identifying attribute
Only used if a Process by Regions layer is chosen. The field and its values will be added to all features output by the tool.
Transfer all attributes from the region layer over to the output candidate EAs
Only used if a Process by Regions layer is chosen. Each
Load
Save
Run
Close


preEA Tool v1.4.1
Inputs
Merge Options
Outputs
Enable Merging
Seed and merge choices
Seed type:
Global best
Merge into the same polygon until no more options are available
Thresholds (0= disabled)
Maximum population of each output tract:
1000.00
Maximum area of each output tract:
9000000.00
Minimum boundary length (%) for a valid merge:
2.00
Weighting coefficients (0-100)
Population
1
Area
0
Shape factor
2
MERGING PARAMETERS
Enable merging
If this is unchecked then no merging will occur and the tool will only generate and populate the building blocks
Seed Type
Specify how the tool selects the next building block to consider when merging
Maximum population of each output tract
Building blocks will be merged with neighbours until they reach this population size. Setting to zero will disable this constraint
Maximum area of each output tract
Building blocks will be merged with neighbours until they reach this area, measured in map units. Setting to zero will disable this constraint
Minimum common boundary length (%)
Building blocks will not be merged with a neighbour if the common boundary length is less than the specified % of the larger polygon's perimeter. Setting to zero will disable this constraint
Population weighting coefficient
Sets the relative importance of population in determining the order in which building blocks are merged. Merges giving the smallest increase in population are preferred. Set to zero to not consider population.
Area weighting coefficient
Sets the relative importance of area in determining the order in which building blocks are merged. Merges giving the smallest increase in area are preferred. Set to zero to not consider area.
Load
Save
Run
Close

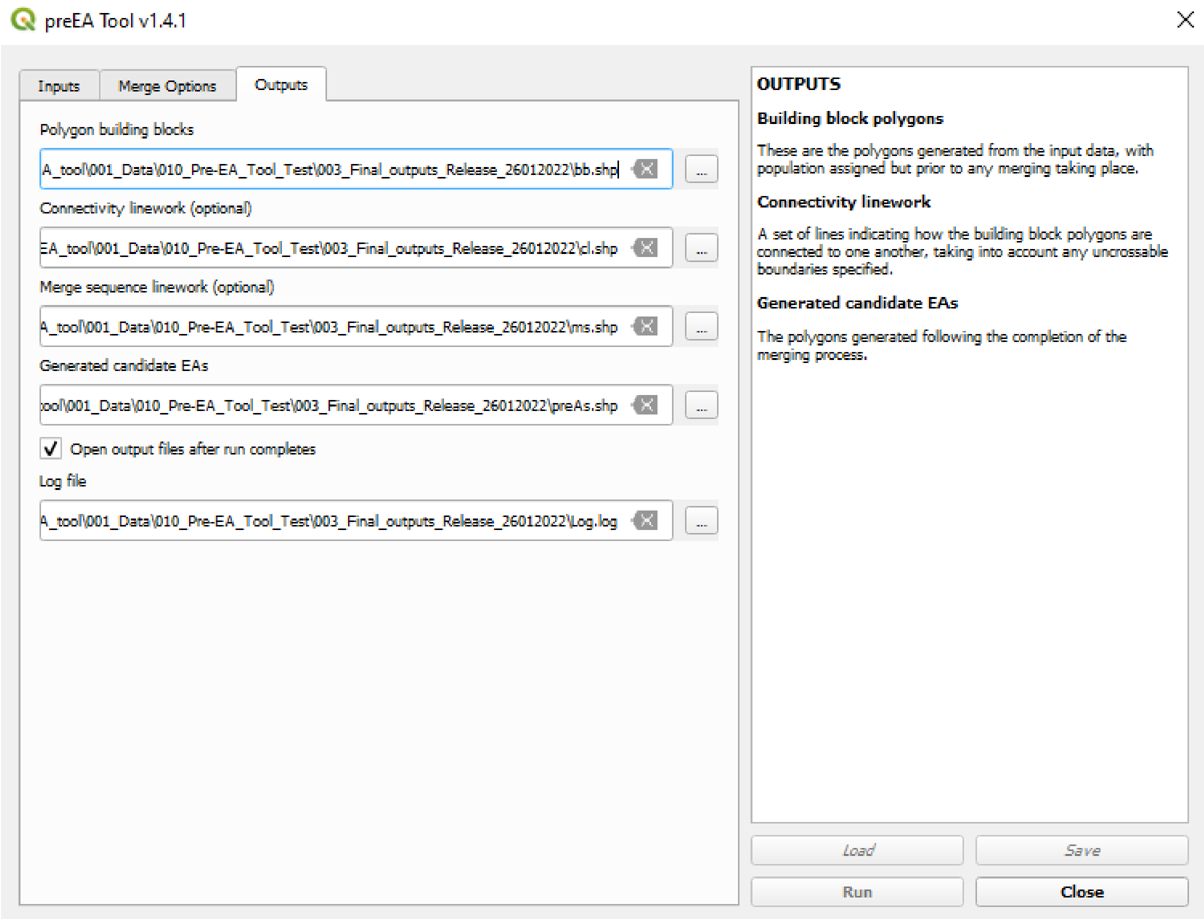


Figure 5. The preEA tool interface with the BFA application settings

3. Results

The 351 Commune administrative boundaries were set as regional constraints for BFA in the tool and all the outcomes from splitting, connectivity linework and preEAs are nested within these commune boundaries. The maximum population size was set to 1000 people and geographic area to 9km^2 in both urban and rural areas since urban/rural classification was not available. Following the splitting process, at the country level, 1,063,329 small polygons (building blocks) were created and out of these units, 61,056 optimal preEAs were generated through the merging process.

The following modifications (i.e. post-processing) were applied to the merged polygons after the tool application:

Several constraints were introduced in the tool to assure that generated preEAs are robust and in line with the country's enumeration area specifications due to limitations caused by the used input data. However, In some areas, generated preEAs are larger than the maximum population and area constraints because these building block polygons were already exceeded one of these maximum constrains due to lack of digitized boundaries. To improve the output, small number of the preEAs with large population and odd shapes were modified manually. In a few other cases, the generated preEAs are smaller, because the merging process stopped due to not finding better solution based on the neighbouring splitting units. To improve these preEAs, a condition to merge preEAs containing less than 100 people and smaller than 1 km^2 was applied to the dataset. To further improve this, a condition to merge preEA containing equal and less than 20 people was applied to the dataset. Both conditions were done automatically using the *Eliminate EAs Meeting Criteria Tool* which has been developed as a postprocessing tool withing the preEA tool package. This tool is designed to be run on the preEA polygon outputs of the preEA tool. It will remove polygons that are too small or otherwise undesirable by merging them with a neighbour with the longest common boundary. The merging process will respect administrative boundaries. The values of the preEA_Popn and preEA_Area variables created by the preEA tool will be summed for the merged polygons. The purpose of applying such conditions is that often census team would prefer to have EAs with certain population size regardless of their spatial coverage. In addition, keeping the EAs with low population will increase the number of EAs which then need more resources during the census cartography. After these processes, the number of preEAs were reduced to 44,570.

We recommend that the preEA outputs should be carefully reviewed in the lab and manually edited as needed prior to census cartography perhaps by using the split polygons (building blocks) that are also provided. Then whilst in the field, the preEA boundaries should be validated and finalised.

Please contact us If the preEA outputs is not matching the criteria that you are looking for in terms of population and area constraints.

Figure 6. Example out, where preEA outlines overlaid on high resolution satellite imagery.

3.1 Release Content and File Descriptions

- **BFA_BuildingBlocks_v1_0.shp**

This is shape boundaries for the building block outputs generated from PreEA tool. The attribute table for the building blocks contains the commune's name (*Nom*), the ISO-3 country code (*Country*), the unique building block IDs (*preEA_BBID*), the estimated total population size (*preEA_Popn*) and the surface area (m^2) (*preEA_Area*). The projection is the projected coordinate system WGS84 (WGS_1984_UTM_Zone_30N).

- **BFA_preEAs_v1_0.shp**

This is shape file for the preEnumeration Area (preEAs) outputs generated from the PreEA tool. The attribute table for the preEAs contains the commune name (*Nom*), the ISO-3 country code (*Country*), the unique preEA IDs (*preEA_EAID*), the estimated total population size (*preEA_Popn*) and the surface area (m^2) (*preEA_Area*). The projection is the projected coordinate system WGS84 (WGS_1984_UTM_Zone_30N).

4. Discussion and Limitations

This work represents the first attempt to create automatic preEAs in BFA where a digital EA dataset from previous census is not in existence. The preEA dataset accompanied by splitting units (building blocks) were successfully generated for entire BFA. The aim of this work was to support the National Institute of Statistics and Demography (INSD) in digitalisation by semi-automatically creating preEnumeration Area datasets that can speed up and make the digital EAs demarcation more robust.

In terms of the input data used in producing the preEAs, IGB and OSM were used as the primary data sources for roads, railways and waterways. The lack of spatial coverage of these datasets has limited the flexibility of the creation of optimal preEAs resulting in some cases with too large or too small preEAs. However, the *Eliminate EAs Meeting Criteria postprocessing tool can enable users to easily adjust these.* It should be noted that due to the nature of the input datasets, odd shape preEAs might still exist in the output and if such boundaries create difficulties for the enumerators. The preEAs needs to be assessed and if needed be modified in the lab prior to the fieldwork.

Even though the high-resolution gridded population dataset used for this work was built on the most recent census data, the total population in each preEA is still an estimate and it might not 100% correspond to the true population on the ground. This is because the census data was only released at admin 3 level, and these were disaggregated statistically (WorldPop and Institut National de la Statistique et de la Démographie du Burkina Faso. 2020).

Acknowledgement

Funding for the development and dissemination of this data set was provided by the Bill & Melinda Gates Foundation and the United Kingdom's Foreign, Commonwealth & Development Office (OPP1182425).

Sarchil Qader supported the generation of inputs for the application of automatic preEnumeration Areas. He was also responsible for data cleaning, processing and

applying the tool to generate final preEAs for BFA. We would like to sincerely thank Mathias Kuepie and Edith Darin for their excellent engagement with BFA team and their constant feedback on the tool and outputs. The high-resolution population dataset was created by Edith Darin. The digitized waterway lines (HYD_Course_eau.shp), National Commune administrative boundary and railroad (Voie_railroads.shp) data set were provided by Geographic Institute of Burkina Faso (IGB). Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org".
We thank the reviewers (Attila Lazar, Edith Darin and Heather Chamberlain) for their careful reading of the report and their constrictive remarks. The whole WorldPop group and GRID3 partners are acknowledged for overall project support.

License

Suggested Citation

Qader S H, Harfoot A, Kuepie M, Darin E, Juran S, Lazar AN, Tatem AJ. 2022. National automatic preEnumeration Areas (preEAs) in Burkina Faso (2019), version 1.0. WorldPop, University of Southampton. Doi:10.5258/SOTON/WP00731

References

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.

Institut National de la Statistique et de la Démographie (2019). “Recensement Général de La Population et de L’habitation de 2019 Du Burkina Faso - Résultats Provisoires.” INSD Ouagadougou, Burkina Faso.