**Data inputs and method to create pre-census Enumeration Areas for three test sites in DRC**

## Summary

**Background:** The Democratic Republic of the Congo (DRC) are preparing for the 2nd National Population and Housing Census. The first Census was conducted in 1984 and with several decades having elapsed since then, there is currently no national enumeration area (EA) dataset in existence for DRC. National censuses typically use EAs as a way to divide areas into small homogeneous units, to ensure that complete household and population enumeration takes place during the census. As part of the GRID3 workshop on modelled population estimates, which took place in Kinshasa June 25th – 28th 2019, WPFM presented an automatic tool which could be used to create small units bounded by natural and manmade features that are visible on the ground (e.g. roads). Automatically creating such units has the potential to facilitate improved logistical planning for the Census Cartography, which also provides an opportunity to validate the unit boundaries in the field. In this report, these units are referred to as pre-census Enumeration Areas (EAs) - a translation of the terminology used by staff from the Bureau Central de Recensement (pre-AD/pre-Aires de Dénombrement).

**Purpose:**
Provide details on the data inputs and steps to generate pre-census EAs for three sites that are chosen for the field test.

**Field Sites:**
Three field sites were selected by BCR to generate pre-census EAs and conduct the field assessment.

- Site 1: Q. KINGU, Commune de Selembao, Kinshasa (Urbain)
- Site 2: Q. DUMI, Commune de Maluku, Kinshasa (Sub-urbain)
- Site 3: Secteur KASANGULU, Territoire Kasangulu, Kongo-Central (Rural)

## Part 1: Data Inputs

1. High gridded population estimates. WorldPop has currently developed a bottom up population modelling to predict population at 100m for five provinces in DRC.
2. Road Data. The road datasets were obtained from Open Street Map (OSM). In addition, BCR has provided some road data in Knigu which was digitized in 2017.
3. Waterway data. The Waterway datasets were obtained from OSM.
4. Rail way data. The rail way datasets were obtained from OSM.
5. Building footprints.
6. Secteur boundary (Admin 3). BCR has provided the dataset.

## Part 2: Methodology

*Step 1: Data preparation*

1- Converts population raster datasets to polygon features. This step helps identify the outline of the settled areas.

2- Euclidean Allocation. This approach calculates, for each cell, the nearest source 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 (settled areas) by using polygon boundaries. The created polygon boundaries are not cutting through the settled areas.

3- Resample the 100m2 population estimates to 10m$^2$. After splitting the Area of Interest (AOI), sometimes, the resulted units are much smaller than 100m$^2$. Therefore, resampling the population data to 10m$^2$ helps allocate people to those small units.

*Step 2: Split*

The aim of the splitting process is to partition the AOIs into regions that are as small as possible so that the subsequent merging process has enough flexibility to combine them into optimal pre-census EAs.

1- The AOIs were split using road, railway, Euclidian Allocation delineation boundaries, waterway and administrative boundaries, using the feature to polygons tool in ArcGIS. These datasets where either lines or polygons whose geometry will be used to small area features. This step results in a set of fully contiguous units for each site with no gaps.

2- After splitting process, we were able to compute the population size for each small unit using the high-resolution population datasets in the Zonal Statistics Tool in ArcGIS.

3- It should be noted that sometimes the resulted splitting units may be larger than defined population and area constraints mainly due to lack of data. Therefore, additional techniques should be employed to split further the large units.

4- For the purpose of preventing pre-census EAs boundaries crossing the river and waterway, the final splitting outcome was subtracted to the water way.

*Step 3: Merging*

When all split regions have population size and area, the regions are then merged until they match constraints using a python script that was developed by WPFM. The maximum population was set to 1200 people in urban, 1000 people in rural and area to 9 million Km$^2$. The script uses the population, area and compactness constraints to merge the small regions.

**Initial results of pre-census EAs in the selected sites:**

**Site one:** Q. KINGU, Commune de Selembao, Kinshasa


Figure 1. Manually digitized road by BCR in 2017.


Figure 2. Road data from OSM.


Figure 3. Splitting the area based on road data

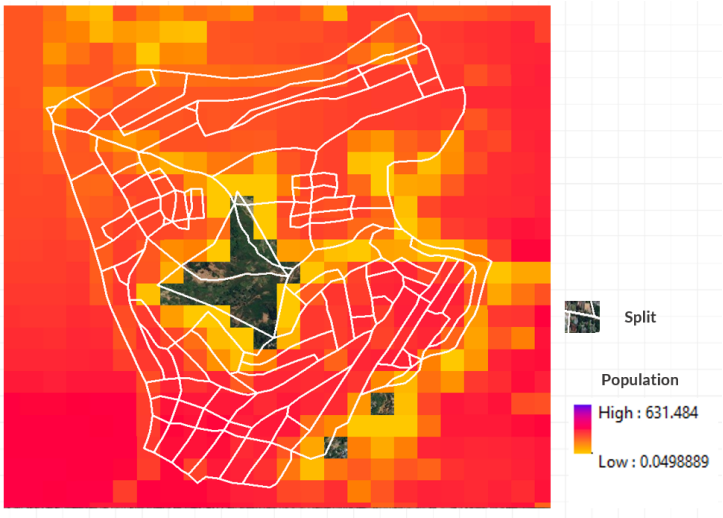


Figure 4. Split regions are overlaid on high gridded population datasets.

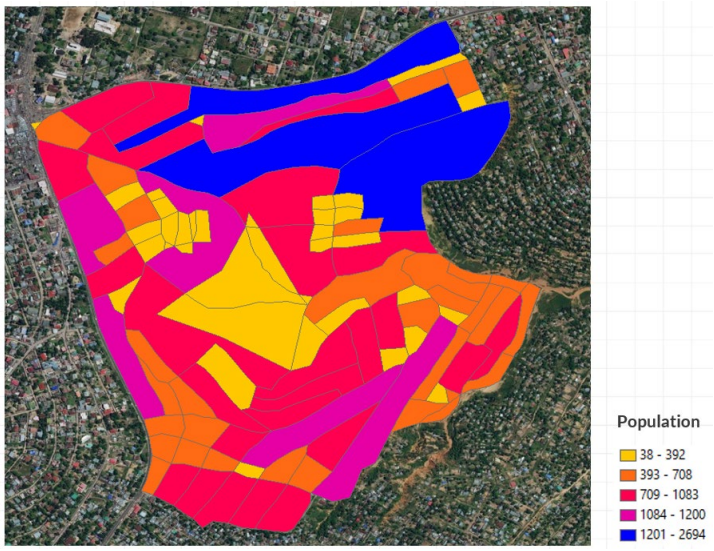

Figure 5. Population is computed for each split unit.

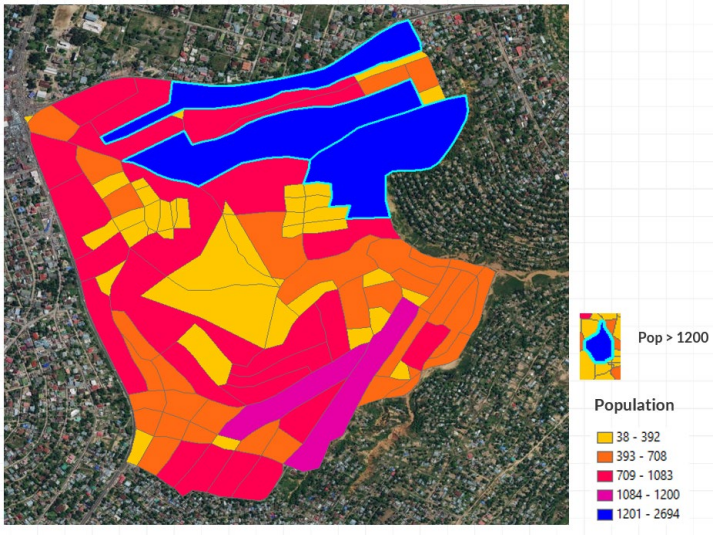

Figure 6. Three pre-EAs are exceeding the maximum threshold (1200 people).

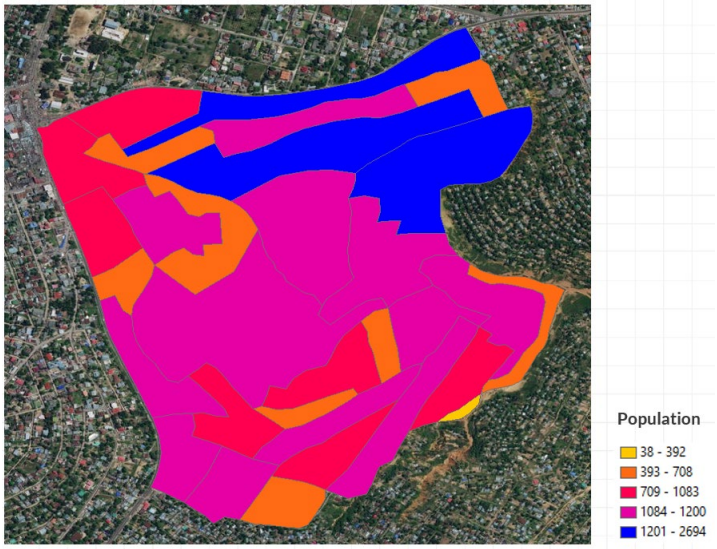

Figure 7. Shows the merged pre-census EAs using 1200 people and 9 million km$^2$ as the maximum constraints for population and area in Kingu.

**Site tow:** Q. DUMI, Commune de Maluku, Kinshasa

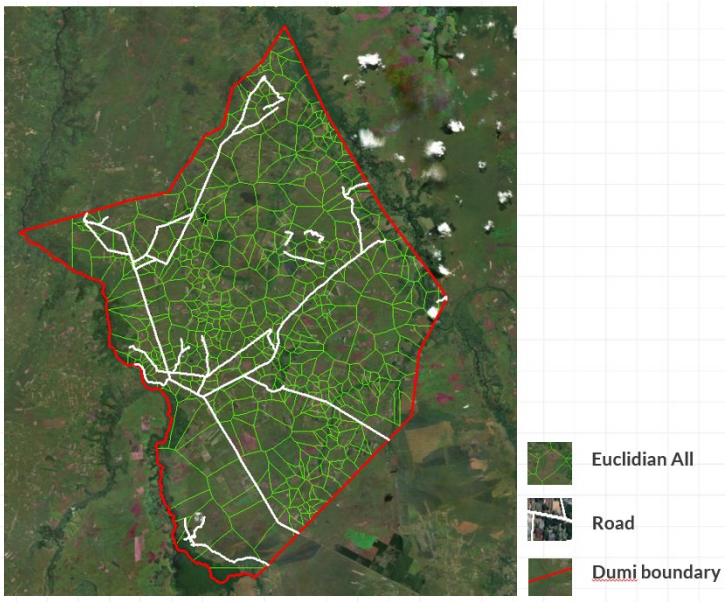


Figure 8. Split Dumi using road, Euclidean Allocation delineation and Dumi boundary.

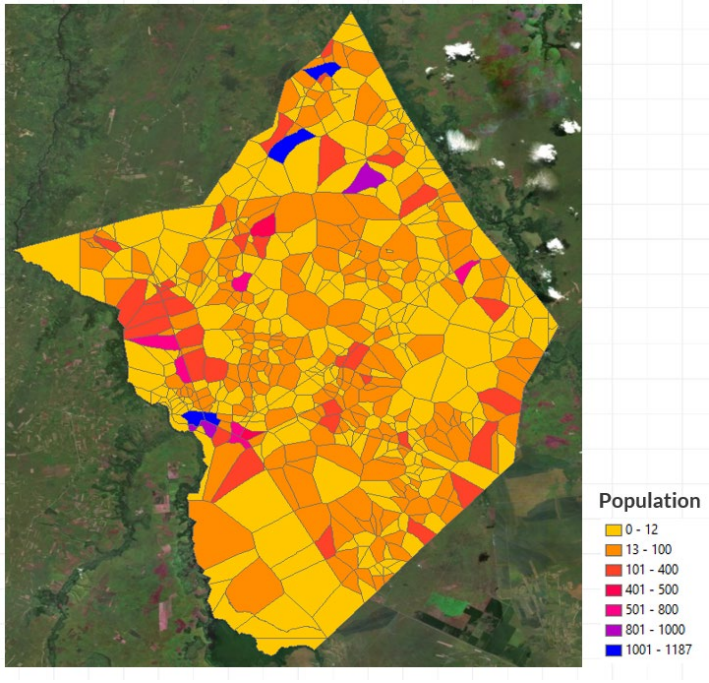


Figure 9. Population is computed for each split unit.

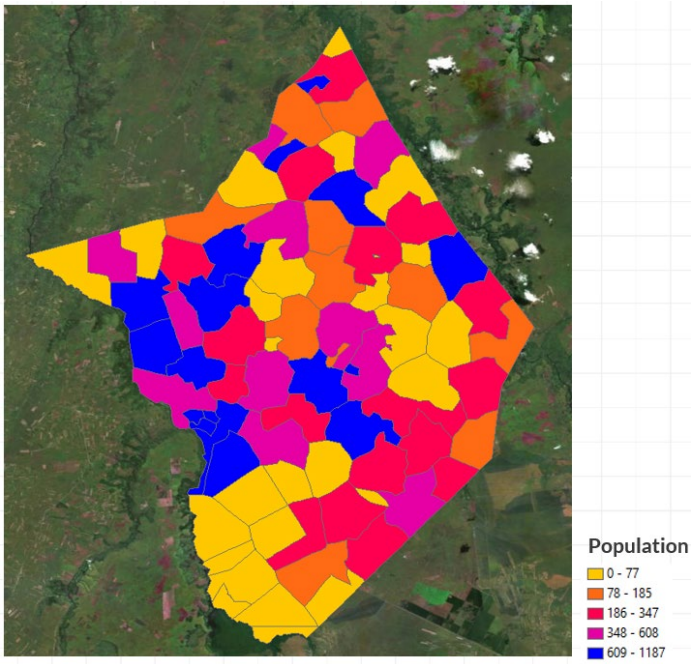


Figure 10. Shows the merged pre-census EAs using 1000 people and 9 million km$^2$ as the maximum constraints for population and area in Dumi.

**Site three:** Secteur KASANGULU, Territoire Kasangulu, Kongo-Central

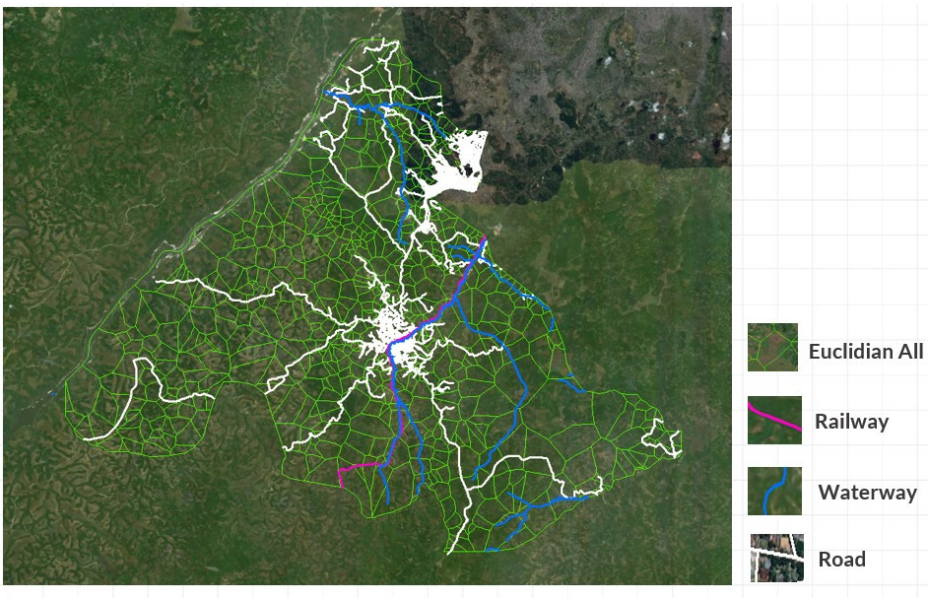


Figure 11. Available road, Railway, Euclidean Allocation delineation and waterway boundary in Kasangulu.

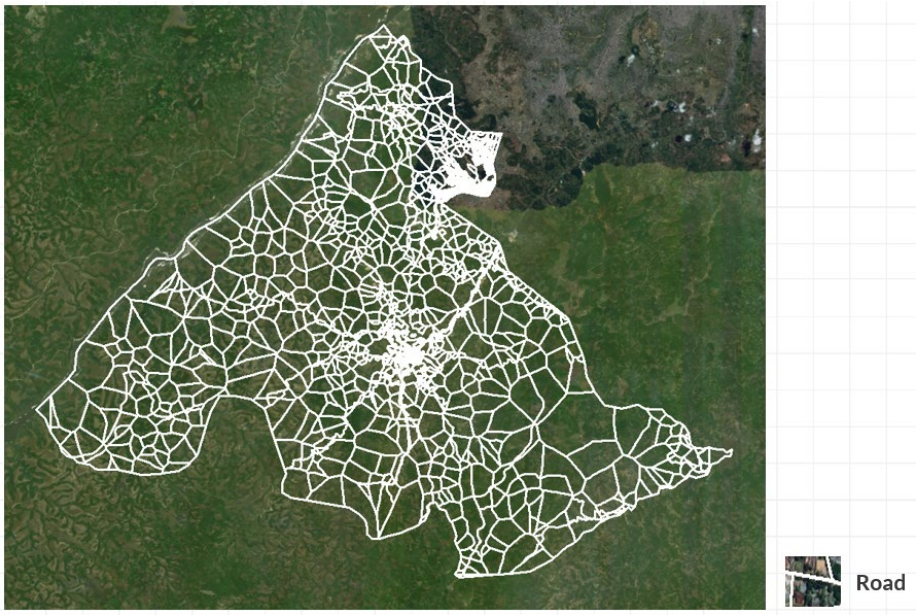


Figure 12. Kasangulu was split using road, Railway, Euclidean Allocation delineation, waterway and Dumi boundary into small units.

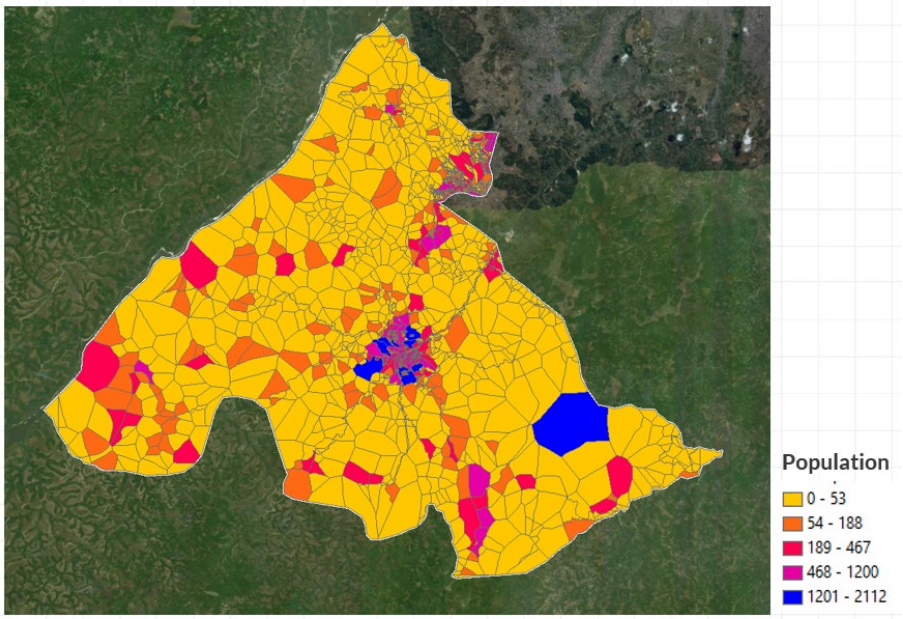


Figure 13. Population is computed for each split unit.

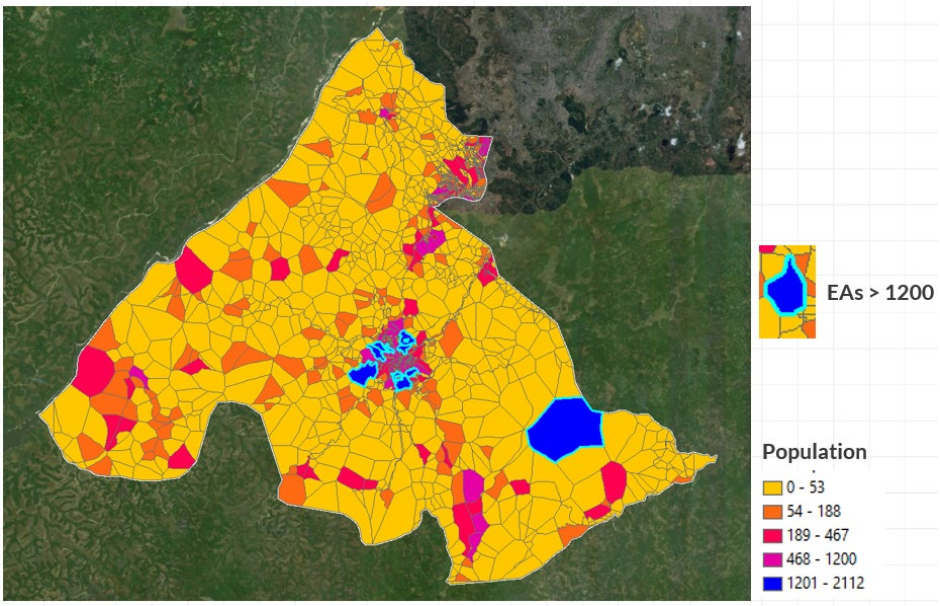

Figure 14. Nine pre-census EAs are exceeding the maximum population size (1200).

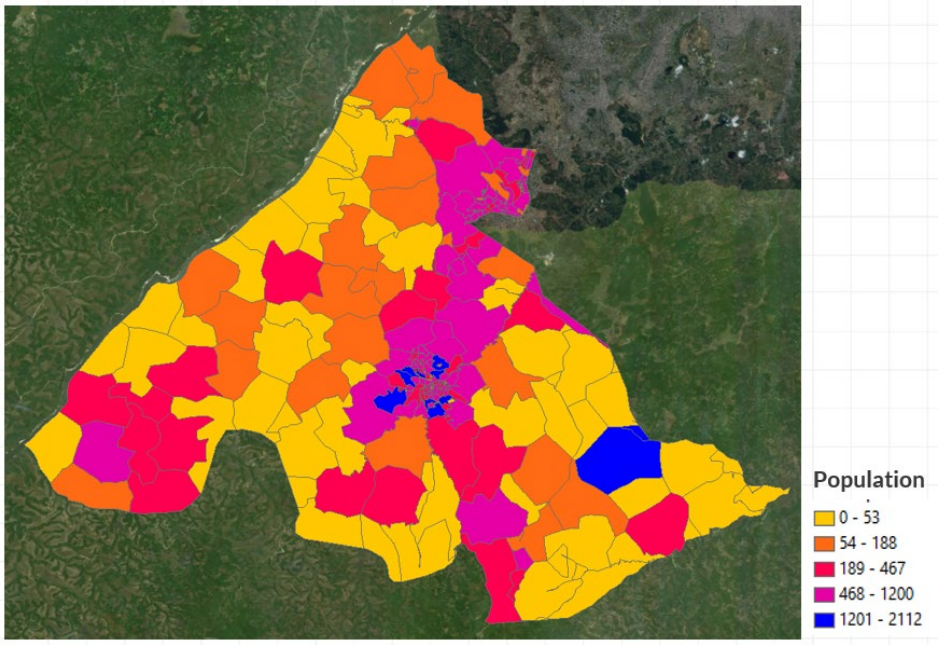

Figure 15. Shows the merged pre-census EAs using 1000 people and 9 million km$^2$ as the maximum constraints for population and area in Kasangulu.

**Further splitting pre-census EAs that are exceeding the maximum population size (1200)**

## Suggested Citation

Qader S.H., Chamberlain H.R., Kuepie M., Hunt F.K., Lazar A.N., and Tatem A.J. 2023. Field testing of pre-Enumeration Areas created using semi-automated delineation approach, Democratic Republic of Congo. WorldPop, University of Southampton. Doi: 10.5258/SOTON/WP00759